VARIATIONAL IMAGE PROCESSING ALGORITHMS FOR THE STEREOSCOPIC SPACE-TIME RECONSTRUCTION OF WATER WAVES

A Thesis
Presented to
The Academic Faculty

by

Guillermo Gallego Bonet

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
May 2011

VARIATIONAL IMAGE PROCESSING ALGORITHMS FOR THE STEREOSCOPIC SPACE-TIME RECONSTRUCTION OF WATER WAVES

Approved by:

Professor Patricio Vela, Committee Chair
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Anthony Yezzi, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Allen Tannenbaum
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Francesco Fedele
School of Civil and Environmental Engineering
Georgia Institute of Technology

Professor Magnus Egerstedt
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Frank Dellaert
School of Interactive Computing
Georgia Institute of Technology

Date Approved: November 22, 2010

To the Father, Esther and my family, who made it possible.

ACKNOWLEDGEMENTS

There are many people that I would like to recognize for supporting me throughout these enriching years at the Georgia Institute of Technology and for participating in this thesis. First, before I pursued doctorate studies in the U.S.A. I was unaware of the power of the Calculus of Variations and its broad applicability to many problems in Electrical Engineering. It is under the guidance of my advisor, Professor Yezzi, that I became to appreciate and embrace this topic in the context of Control Theory applied to Image Processing and Computer Vision. I also thank my co-advisor, Professor Fedele, for introducing me into the topic of Ocean Engineering, the central application of this thesis. I am very grateful to both my advisors for mentoring and supporting me in my research, for their creativity and fruitful discussions that helped me develop this thesis. I am indebted to Dr. Benetazzo from Protecno S.r.l. for his collaborations in this project and to Dr. Ardhuin for the Crimea data shared to carry out the experiments. I thank Professors Tannenbaum, Vela, Egerstedt, Dellaert and Wolf for serving in my proposal and/or dissertation defense committees.

It has been a pleasure interacting with other members of the Laboratory of Computational Computer Vision (LCCV), past and present. I wish to thank the people and collaborators that I have met during my stays in CIRAM and CINECA (Bologna), and in Telefonica I+D (Barcelona). I am indebted to the staff at the LCCV, the Graduate Office, the Office of International Education (OIE), and the Institute of International Education (IIE) for helping me dedicate my time on research instead of paperwork. I would also like to acknowledge the Fulbright commission (Spanish program) for providing me the opportunity to pursue graduate studies in the U.S.A.

I am grateful to my friends, in the U.S.A. and in Europe, who have walked with me along this difficult journey. Last but not least, I honor my family, and the ones that are to join and to come, for their everlasting love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

A novel video observational method for the space-time stereoscopic reconstruction of dynamic surfaces representable as graphs, such as ocean waves, is developed. Variational optimization algorithms combining image processing, computer vision and partial differential equations are designed to address the problem of the recovery of the shape of an object's surface from sequences of synchronized multi-view images. Several theoretical and numerical paths are discussed to solve the problem. The variational stereo method developed in this thesis has several advantages over existing 3-D reconstruction algorithms. Our method follows a top-down approach or object-centered philosophy in which an explicit model of the target object in the scene is devised and then related to image measurements. The key advantages of our method are the coherence (smoothness) of the reconstructed surface caused by a coherent object-centered design, the robustness to noise due to a generative model of the observed images, the ability to handle surfaces with smooth textures where other methods typically fail to provide a solution, and the higher resolution achieved due to a suitable graph representation of the object's surface. The method provides competitive results with respect to existing variational reconstruction algorithms. However, our method is based upon a simplified but complete physical model of the scene that allows the reconstruction process to include physical properties of the object's surface that are otherwise difficult to take into account with existing reconstruction algorithms. Some initial steps are taken toward incorporating the physics of ocean waves in the stereo reconstruction process. The developed method is applied to empirical data of ocean waves collected at an off-shore oceanographic platform located off the coast of Crimea, Ukraine. An empirically-based physical model founded upon current ocean engineering standards is used to validate the results. Our findings suggest that this remote sensing observational method has a broad impact on off-shore engineering to enrich the understanding of sea states, enabling improved

design of off-shore structures. The exploration of ways to incorporate dynamical properties, such as the wave equation, in the reconstruction process is discussed for future research.

CHAPTER I

INTRODUCTION

Wind-generated waves play a prominent role at the interfaces of the ocean with the atmosphere, land and solid Earth. Waves also define in many ways the appearance of the ocean seen by remote-sensing instruments. Classical observational methods for ocean waves rely on time series retrieved from wave gauges and ultrasonic instruments or buoys to measure the space-time dynamics of oceanic sea states. Global altimeters, or Synthetic Aperture Radar (SAR) instruments are exploited for observations of large oceanic areas via satellites, but details on small-scales are lost. To monitor oceanic areas with a wavenumber spectral resolution that ranges from $100m^{-1}$ to $0.01m^{-1}$, video-based observational methods relying on computer vision techniques have recently appeared [4, 107]. Such systems use two or more stereo camera views pointing at the ocean to provide spatio-temporal data and statistical content richer than that of previous monitoring methods. Vision systems are non-intrusive and have economical advantages over their predecessors, but they require more processing power to extract information from the ocean.

In this thesis, a novel video observational method based on variational stereo techniques is proposed to obtain accurate estimates of the spatio-temporal dynamics of ocean waves for the spectral range previously mentioned. The thesis focuses on the computer vision component of the broader ocean engineering problem: the recovery of the spatio-temporal shape and radiance of ocean waves given stereo videos acquired by calibrated cameras in off-shore structures. New image processing and vision algorithms based on partial differential equations (PDEs) are presented in the context of active surfaces to obtain a feasible solution to the problem.

This solution to the vision problem will be used in a larger system called the Variational Wave Acquisition Stereo System (VWASS) to monitor the statistics of the water climate around off-shore structures to predict and to prevent natural disasters (floods) or to decide

the optimal location to build off-shore platforms. The algorithm will be installed at the Venice Lagoon as part of an early warning system to detect high tide so that flooding of the city can be prevented by raising protecting gates as part of the Modulo Sperimentale Elettromeccanico[1] (MOSE) project.

1.1 Organizational structure of the thesis

This thesis is organized as displayed in Figure 1. Chapter 1 introduces the problem that motivates the research in stereo methods for dynamic surfaces. This problem is exemplified by an ocean engineering application: the measurement of ocean waves. Previous literature on the topic is reviewed both from ocean engineering and computer vision points of view. Research objectives are stated and contributions of this thesis are presented. In chapter 2, a background review on classical stereo methods for still images is given. Motivated to overcome the shortcomings of previous stereo methods, variational stereo methods are also introduced.

Chapter 3 presents the novel variational method for the three-dimensional reconstruction of surfaces that admit a representation in the form of a graph. Advantages and disadvantages with respect to existing object-based variational stereo methods is given prior to its mathematical development. The reconstruction problem is cast into a variational optimization problem in the context of active surfaces. The numerical solution of the system of PDEs that characterizes the solution of the optimization problem is discussed, and multigrid methods are justified as the most efficient PDE solvers. Examples of the application of the proposed variational stereo method to still images are also given. Several topics in this and other chapters have been moved to the appendices to maintain focus of the thesis.

In chapter 4, an object-based photometric stereo method to estimate the gravity plane that supports the graph of ocean waves is discussed, both theoretically and numerically. Chapter 5 takes a step in the direction of incorporating physics of the surface in the three-dimensional estimation process. Geometric and statistical regularizers are presented in the same variational framework of previous chapters. To overcome scenarios with sea foam

[1] in English, Experimental Electro-mechanical Module

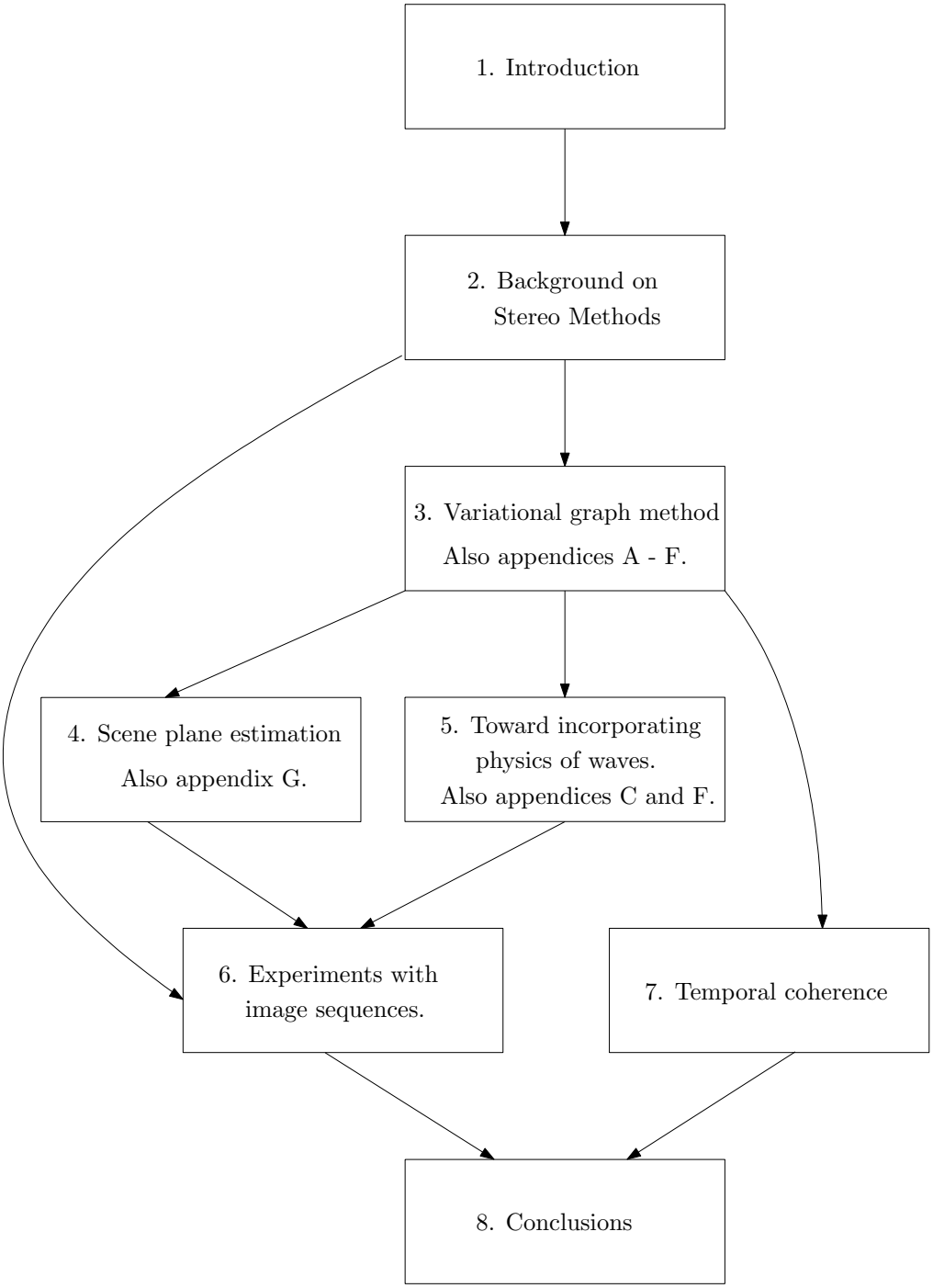


Figure 1: Structure of the thesis, according to chapters.

and regions of specular reflections, a modification of the photometric criterion used in the data fidelity component of the variational stereo method is given. Chapter 6 extends the variational stereo method developed in chapter 3 for still images to the case of a sequence

of images. Chapter 6 also presents the validation of the stereo method on empirical data, both from image processing and ocean engineering points of view. Supporting evidence of the competitive performance of the algorithm with respect to other methods is reported.

Chapter 7 goes beyond sequential processing of stereo images to propose the incorporation of temporal coherence of the reconstructed dynamic surface by casting the reconstruction problem as the estimation of a 3-D manifold of graphs in a 4-D space. This implies the simultaneous estimation of the surface for all stereo images in the sequence. Conclusions and final comments on future research directions are discussed in chapter 8.

1.2 Origin and history of the problem

Since this thesis covers both the topics of shape reconstruction and oceanic sea states, it relates to a vast body of literature. Stereography is a classical remote sensing technique that was developed in the early 1990s to infer the depth of opaque surfaces from images. Water surface topography can be measured from conventional stereographic algorithms [59], which are used to survey geodetical surfaces or static objects. The major difference in our case is that the water surface is a specular object in rapid movement. Hence, only a spatio-temporal description can correctly characterize a changing object like the water surface. The recovery of the spatio-temporal structure of water waves from stereo videos is possible by adding a temporal dimension to classical stereographic algorithms.

The first experiments with stereo cameras mounted on a ship were by Schumacher [84] in 1939. Later, Coté et al. [13], in 1960, demonstrated the use of stereophotography to measure the sea topography for long ocean waves from airbone platforms in the North Atlantic Ocean during the Stereo Wave Observation Project (SWOP). The study of long waves using stereophotography was also discussed by Sugimori [97], based on an optical method by Barber [67], and by Holthuijsen [42]. Since extracting a 3-D model from an image pair is labor intensive compared to conventional wave height measuring techniques, it was not until significant advances in computers were made in the 1980s that this technique gained popularity in the study of the dynamics of oceanographic phenomena.

Nevertheless, the need to know directional information for waves allowed stereography to

remain as a useful tool in oceanographic studies. In this context, Shemdin et al. [91, 89, 90] applied stereography for the directional measurement of short ocean waves, with wavelengths in the range of 1-100 cm. During the Tower Ocean Wave and Radar Dependence (TOWARD) experiment, Shemdin et al. used a pair of cameras mounted on an oceanographic off-shore tower near San Diego, California, to create a 3-D model of the sea surface, and then they extracted directional information of waves via spectral analysis. They also performed experiments from another platform near Cape Henry, Virginia [100]. In this, they concluded that stereo-photography is a viable method for measuring the directional wave number spectra of short gravity waves. In 1997, Holland et al. [41] demonstrated the practical use of video systems to measure nearshore physical processes. A more recent integration (2004) of stereographic and photogrammetric techniques into the field of oceanography has been the WaveScan project (Santel et al. [81]), as a collaboration between two Institutes within the University of Hannover (Germany). Given images of nearshore oceanic events, Santel et al. generated a three-dimensional point cloud via feature matching and then interpolated a surface model that represents the ocean surface.

The reconstruction of the ocean wave surface from stereo pairs of images is an example of a classical problem in computer vision commonly known as the correspondence problem [59, 35]. Its solution is based upon epipolar geometry techniques that find corresponding points in pairs of images, from which one is able to estimate the real point in the 3-D terrestrial coordinate system by triangulation [38, 35]. Benetazzo [4] successfully incorporated these techniques in the Wave Acquisition Stereo System (WASS) [4] and tested it in experiments off the shore of the California Coast (San Diego) and the Venice coast in Italy. He was able to estimate wave spectra from the extracted time series of the surface fluctuations at one fixed point from the data images. The accuracy of such spectral estimates is comparable to spectral estimates obtained from ultrasonic transducer measurements. An example of a WASS system currently installed in the Acqua Alta platform is shown in Figure 2.

An alternative trinocular imaging system (ATSIS) for measuring the temporal evolution of 3-D surface waves was proposed in [107]. This trinocular system uses three cameras (in a configuration of three pairs of stereo cameras: left-right, left-center, and center-right) to

Figure 2: Left: Off-shore platform "Acqua Alta" in the Northern Adriatic Sea, near Venice. Center: Pair of synchronized cameras for monitoring the ocean climate from the platform. Right: WASS hardware installed at the platform for recording stereo videos of ocean waves.

exploit the consistency constraints among the views such that a more accurate reconstruction of the water surface can be obtained with classical stereo algorithms than over that of a single pair. More recently, in [29] it is shown how a modern variational stereo reconstruction technique pioneered by Faugeras and Keriven [19] can be applied to the estimation of oceanic sea states. Additional references demonstrate that this is an active research topic in the ocean engineering community [61, 95, 34, 55, 44, 60].

The three-dimensional shape reconstruction of an object's surface from stereo pairs of images is also a classical problem in computer vision (see, for example [104, 35, 18, 88, 25, 59] and references therein), and it is still an extremely active research area. Recent surveys of 3-D reconstruction algorithms are reported in [58, 85]. There are many 3-D reconstruction algorithms available in the literature and the reconstruction problem is far from being solved. The different algorithms are designed under different assumptions and provide a variety of trade-offs between speed, accuracy and viability. Embedded in the design, they can significantly differ in the sparsity of their input and output, e.g., in the way visual cues are extracted from data images and the way the 3-D scene being reconstructed is represented. For example, some algorithms provide a scattered representation of the surface of the objects, while others output a dense reconstruction.

Since there are many available algorithms, the selection of the appropriate one is heavily

dependent on the application and the available resources. Lu et al. [58] propose a first natural classification of reconstruction algorithms according to whether an a priori calibration of the cameras is given or it can be obtained at run time. This distinction is motivated by practical considerations. We put forward that, since we are interested in obtaining a 3-D model of the ocean surface as accurately as possible, it is crucial to have a pre-calibrated camera acquisition system. Algorithms within both major categories (*pre-calibrated* vs. *on-line calibrated*) can be further classified according to other characteristic hypotheses that determine their design. Within the pre-calibrated class, *image-based* algorithms that obtain reconstructions from either sparsely matched features or dense correspondences are the most popular ones. The graph variational method in this thesis falls in the category of *object-based* algorithms that use variational principles to deform an initial surface toward the object to be reconstructed. The third sub-class, *voxel-based* algorithms, corresponds to those that work by discretizing the scene into a set of voxels and the reconstruction of the scene is obtained by shape carving [86] or by optimization methods [106]. The on-line calibrated class is further categorized according to how the calibration of the cameras is estimated: via a priori *scene constraints* (taking advantage of known information of parallel lines in the scene, vanishing points [9, 54], etc.) or via *geometric constraints* derived from reasonable assumptions on the the intrinsic parameters of the cameras. The latter is the class of flexible self- or auto-calibration algorithms [62, 37, 102, 72, 79, 75, 78], where geometrical entities, such as the Absolute Conic or the Dual Absolute Quadric, that encapsulate the Euclidean structure of the scene appear.

Traditional *image-based* stereo methods follow a *bottom-up* approach to recover, from elementary image primitives, the surface of a target object in the 3-D world. This approach consists of three steps: first image primitives (points or regions) are detected and matched across images by optimizing a photometric score to establish local correspondences; then depth is inferred by combining these correspondences using *triangulation* of 3-D points (also known as parallax or *back-projection* of image points); finally, surface fitting through the scattered collection of 3-D points returns a continuous model of the real-world surface. The first step, also known as the stereo matching problem, is significantly more difficult than the

second one. However, epipolar geometry between image pairs can be exploited to reduce feature matching to a 1-D search along epipolar lines. This approach has the advantages of being simple and fast. Unfortunately, it also has two major disadvantages that motivated the research on improved stereo reconstruction methods [19, 110, 46]. These disadvantages are: (i) Correspondences rely on strong textures (high contrast between intensities of neighboring points). Image matching gives poor correspondences if the objects in the scene have a smooth radiance (slow varying intensity). Correspondences also suffer from the presence of noise and local minima. (ii) Each space point is reconstructed independently and therefore the recovered surface of an object is obtained as a collection of scattered 3-D points (a *point cloud*). Thus, the hypothesis of the continuity of the surface is not exploited in the reconstruction process. The breakdown of traditional stereo methods in these situations is evidenced by "holes" in the reconstructed surface, which correspond to unmatched image regions [59, 4]. This phenomenon may be dominant in the case of the ocean surface, which, by nature, is generally continuous and contains little texture.

Modern *object-based* image processing and computer vision methods that rely on Calculus of Variations and Partial Differential Equations (PDEs), such as Stereoscopic Segmentation [110] and other variational stereo methods [19, 3, 46], follow a *top-down* approach or *object-centered* philosophy in which an explicit model of the target object in the scene is first devised and then related to image primitives. These methods are able to overcome the disadvantages of traditional stereo methods. For instance, unmatched regions are avoided by building an explicit model of the smooth surface to be estimated rather than representing it as a collection of scattered 3-D points, or by building a smooth disparity map[2] for correspondence establishment rather than a scattered set of matched image points. Thus, variational methods provide dense and coherent (continuous) surface reconstructions. Surface points are not reconstructed independently or continuously along epipolar lines (1-D coherence) as in traditional stereo, but by exploiting the continuity hypothesis in the full two-dimensional domain of the surface or the disparity map. Variational stereo methods

[2] A vector field between two images specifying the difference in horizontal and vertical components of corresponding points.

combine correspondence establishment and shape reconstruction into one single step. They are less sensitive to matching problems of local correspondences than traditional stereo since feature detection and matching have been eliminated. The reconstructed surface is obtained by minimization of an energy functional designed for the stereo problem. This functional measures the photo-consistency of the reconstructed model via an image integral or a weighted area integral. The solution is obtained in the context of active surfaces by deforming an initial surface or disparity map via a PDE in the form of a gradient descent flow, derived from the necessary optimality conditions of the energy functional, the so-called Euler-Lagrange (EL) equations. In contrast with traditional stereo, a back-projection model is not needed in object-based variational stereo since 3-D point triangulation is never explicitly performed during the deformation process; only the knowledge of the forward projection model is necessary.

1.3 Research objectives

The objective of the research in this thesis is the theoretical design of object-based variational stereo methods as well as efficient numerical algorithms for the estimation of the shape of ocean waves observed by means of a stereo video camera system.

From the theoretical point of view, this means building upon existing stereo methods and combining their best properties to successfully reconstruct the object of interest, in this case, oceanic sea states. More specifically, this translates to merging variational 3-D reconstruction [19] with generative image models of smooth scenes, though our goal is not solving the segmentation problem [110]. It also requires the study of new energy functionals to improve the reconstruction results, and the overcoming of the difficulties that arise from the application of newly designed stereo methods to our dynamic and physical application on ocean waves.

To carry out successful research from such a rich and large video data source requires not only a good theoretical model, but also an efficient numerical implementation to obtain results in a reasonable amount of time. In addition, flexibility and scalability of the numerical method are crucial to allow the adoption of new research directions. More specifically,

from a numerical point of view, the main research objective implies the study and adoption of flexible acceleration techniques for systems of partial differential equations and temporal processing of the stereo video data. Among the several acceleration techniques considered are preconditioning, multiresolution and parallel processing.

1.4 Contributions of this thesis

The main contribution of this thesis is the design and implementation of a variational observational method to solve the problem of the spatio-temporal reconstruction of oceanic sea states given stereo video. Variational optimization tools are developed and applied to solve the reconstruction problem in the context of active surfaces. In particular, our method relies on the representation of ocean waves in the form of a graph. A base generative model of the images in the video data is developed and several improvements are also studied.

The variational stereo method developed in this thesis has several advantages over existing 3-D reconstruction algorithms. The key advantages of our object-based method are the coherence (smoothness) of the reconstructed surface caused by a coherent object-centered design, the robustness to image noise due to a generative model of the observed images, the ability to handle surfaces with smooth textures where other methods typically fail to provide a solution, and the higher resolution achieved due to a suitable graph representation of the object's surface. The method provides competitive results with respect to existing variational reconstruction algorithms. However, our method is based upon a simplified but complete physical model of the scene that allows the reconstruction process to include physical properties of the object's surface that are otherwise difficult to take into account with existing reconstruction algorithms. Some steps in this direction are also taken in the thesis. To the authors' knowledge, this is the first time that variational stereo methods have been applied to ocean engineering for the reconstruction of ocean waves.

More specifically, the contributions of this thesis are the following: (*i*) the multi-camera, object-based variational method proposed in chapter 3 for the reconstruction of a generic static surface that admits a representation in the form of a graph; (*ii*) the implementation of a tailored multigrid algorithm discussed in section 3.4 to numerically solve the variational

optimization problem at the heart of the multi-camera method above-mentioned; (*iii*) the extension of the multi-camera snapshot reconstruction method to include temporal processing of consecutive images; (*iv*) the incorporation, using variational principles, of physical properties of the waves in the form of weak constraints (chapter 5); (*v*) the correction of the data-fidelity term to attenuate the effect of specular regions and/or small breaking waves (section 5.3); and (*vi*) the inclusion of temporal coherence in the reconstruction process (chapter 7).

Other smaller contributions that also add value to the thesis are: the inclusion of Automatic Gain Control (AGC) compensation within the generative model of the images, the derivation of the optimality equations for a generic point-wise photometric error function and its particularization for quadratic and sub-quadratic functions, an algorithm to estimate the plane that best explains a quasi-planar oceanic scene by using planar homographies, the analysis of several algorithms to estimate the mean plane through a surface that is represented as a collection of 3-D points, an algorithm to reorient a quasi-planar surface tailored to our application, and the statistical and frequency analysis of the reconstruction results in light of ocean engineering theory.

CHAPTER II

BACKGROUND ON STEREO METHODS

This thesis is strongly motivated by a combination of the existing work in [4] and [110]. In particular, the variational approach of *Stereoscopic Segmentation* [110] is used to tackle the vision problem: the reconstructed surface of the ocean is obtained as the minimizer of an energy functional designed to fit the measurements and dynamics of ocean waves. An explicit smooth model of the surface is obtained as opposed to the time series of scattered 3-D points provided by WASS [4]. We design a novel Variational Wave Acquisition Stereo System (VWASS) based on state-of-the-art stereo reconstruction techniques for accurate estimates of the spatio-temporal dynamics of ocean waves. VWASS processes real data from experiments at offshore platforms (see Figure 2). The stereo video acquired by the setup at the platform constitutes the multi-dimensional input data for our theoretically designed and physically implemented algorithms. The output of VWASS consists of the 4-D (space and time) reconstruction of the sea surface, which may then be used to obtain the statistics of the waves: wave spectra, significant wave height, etc. Preliminary research [29, 20, 4] has shown that VWASS is a promising technology with a broad impact on off-shore engineering since it will enrich the understanding of the statistics of sea states, enabling improved design of off-shore structures.

2.1 Multi-image setup

Let us first analyze the reconstruction of the water surface for a fixed time. The geometry of the image formation process, which states how points in 3-D are mapped onto points in the image plane, is described by the pinhole camera model [104, 35, 18]. In this model, a camera is a device that performs a central projection of the world onto the image plane.

Let $\{I_i\}_{i=1}^{N_c}$ be a set of images of a static (water) scene acquired by N_c cameras whose calibration parameters are $\{\mathtt{P}^i\}_{i=1}^{N_c}$. The i-th camera maps a 3-D point $\mathbf{X} = (X, Y, Z)^\top$ onto the 2-D point $\mathbf{x}_i$ where the *optical ray* from $\mathbf{X}$ (line joining $\mathbf{X}$ and the *optical center*

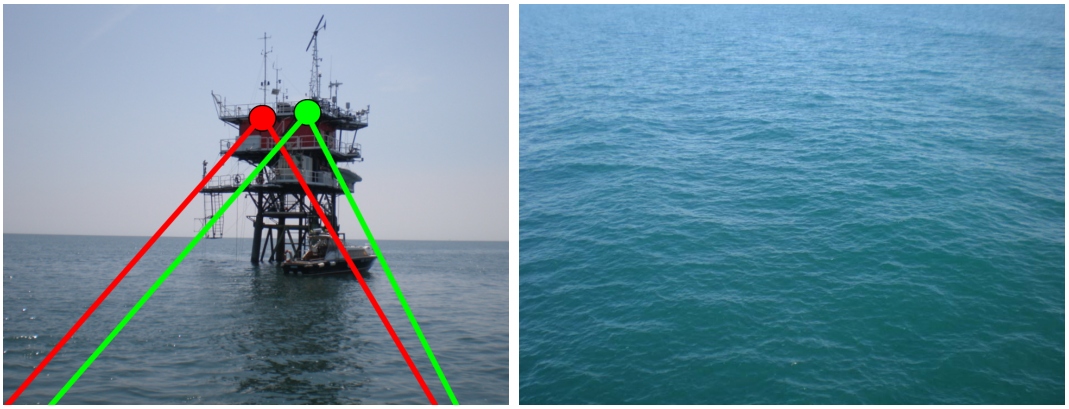


Figure 3: Left: Stereo pair of cameras installed on the off-shore platform and schematic of their field-of-views (FOVs). Right: Sample image of the water surface acquired by one of the cameras.

of the camera[1], $\mathbf{C}_i$) meets the image plane of the camera. The 2-D Euclidean coordinates of this point on the image plane are $\mathbf{x}_i = (x_i, y_i)^\top$. Projective geometry in homogeneous coordinates provides a convenient framework to express such a projection mapping due to the linearity of the equations. In homogeneous coordinates,

$$\bar{\mathbf{x}}_i \sim \mathtt{P}^i \bar{\mathbf{X}}, \tag{1}$$

where $\bar{\mathbf{x}}_i = (x_i, y_i, 1)^\top$, the symbol $\sim$ means equality up to a nonzero scale factor and $\mathtt{P}^i = \mathtt{K}^i[\mathtt{R}^i \mid \mathbf{t}^i]$ is the 3×4 projection matrix with the intrinsic ($\mathtt{K}^i$) and extrinsic ($\mathtt{R}^i, \mathbf{t}^i$) calibration parameters of the i-th camera. The intrinsic parameters are those that specify the affine transformation between camera- and pixel-coordinates within the image plane, and the extrinsic parameters are those that characterize the Euclidean transformation between camera- and world-coordinates in 3-D space. $\mathtt{K}^i$ consists of the focal length of the camera, the coordinates of the principal point where the optical ray perpendicular to the image plane meets the latter, and the shape of the pixels. $\mathtt{R}^i$ and $\mathbf{t}^i$ are the rotation and translation of the camera frame with respect to the world or object coordinate frame. The optical center of the camera is the point $\mathbf{C}_i$ that does not have a defined image projection, i.e.,

$$\mathtt{P}^i \bar{\mathbf{C}}_i = \mathbf{0}. \tag{2}$$

[1] Also known as *center of projection* or *focal point*.

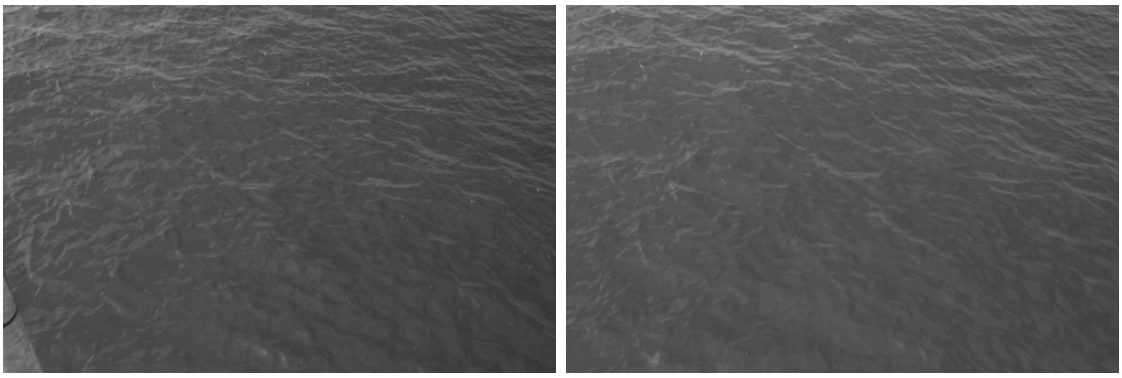


Figure 4: Two images of the surface of the water from an off-shore platform taken at the same time but from different locations, i.e., a stereo image.

Calibrated cameras are those whose intrinsic and extrinsic parameters are known with respect to some world coordinate system, i.e., whose projection matrices are known. In this thesis, all cameras are assumed to be calibrated.

Let $\pi_i : \mathbb{R}^3 \to \mathbb{R}^2$ note the perspective projection maps: $\mathbf{x}_i = \boldsymbol{\pi}_i(\mathbf{X})$. Finally, $I_i(\mathbf{x}_i) \equiv I_i(\pi_i(\mathbf{X}))$ is the intensity at $\mathbf{x}_i$, the coordinates of the projection of $\mathbf{X}$ on the i-th image plane.

2.2 Classical stereo

Following a *bottom-up* point of view, the problem of the 3-D reconstruction of an object's surface from its projections in a pair of images (Figure 4) has the remarkable property of being separable in two subproblems:

1. The establishment of correspondences between image points, $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$.

2. Depth estimation of the corresponding 3-D points.

Solving the first problem (stereo matching) is significantly more difficult than solving the second one. However, the one can take advantage of the so-called *epipolar constraint* to help solve the stereo matching problem.

The projections on two images, $\mathbf{x}_1, \mathbf{x}_2$, of a 3-D point $\mathbf{X}$, satisfy the epipolar constraint

$$\bar{\mathbf{x}}_2^\top \mathtt{F} \bar{\mathbf{x}}_1 = 0, \tag{3}$$

where $\mathtt{F}$ is the 3×3 (homogeneous) *fundamental matrix* between images 1 and 2. This matrix encapsulates, algebraically, the incidence relations on the image planes of a pair of cameras. Assuming $\mathtt{F}$ is known, if an image point $\mathbf{x}_1$ is chosen in image 1, equation (3) states that the corresponding point in the second image, $\mathbf{x}_2$, must lie on the *epipolar line* whose homogeneous coordinates are $\mathbf{l}_2 = \mathtt{F}\bar{\mathbf{x}}_1$. This property is of most significance in searching for a correspondence because it reduces the dimensionality of the search to a 1-D space [74]. In our case of pre-calibrated cameras, the fundamental matrix $\mathtt{F}$ can be computed from the projection matrices of the cameras: $\mathtt{P}^1$ and $\mathtt{P}^2$. Otherwise, the celebrated, normalized 8-point algorithm [36, 35] is the cornerstone toward the estimation of $\mathtt{F}$ from observed point correspondences.

Assuming calibrated cameras, once a correspondence is established, it is straightforward to compute the depth of the 3-D point $\mathbf{X}$ from $\mathbf{x}_1, \mathbf{x}_2, \mathtt{P}^1$, and $\mathtt{P}^2$ by a process called *triangulation* [38, 35] (parallax, in astronomy), sometimes also referred to as *back-projection* of image points.

Classical stereo algorithms solve the correspondence problem by selecting a point $\mathbf{x}_1$ and searching for $\mathbf{x}_2$ among the candidate points on the epipolar line (see Figure 5). Matching is established according to the minimization or maximization of some photometric score between the intensities $I_1(\mathbf{x}_1), I_2(\mathbf{x}_2)$, and possibly more intensities at neighboring locations, where $\mathbf{x}_2$ varies along the epipolar line. A typical choice is the maximization of the cross correlation of the intensities of windows centered at $\mathbf{x}_1$ and $\mathbf{x}_2$. In this case, a crucial parameter is the selection of the size of the search window. Small windows contain few data points, and thus are very sensitive to image intensity noise, often resulting in false matches. On the other hand, large windows may extend to intensities corresponding to different objects or surfaces, also causing a breakdown of the matching process. To mitigate this problem, the equivalent of an adaptive window size was proposed in [64] using a scale space approach.

The matching and depth-recovery process can be repeated for each point independently or for neighboring points along epipolar lines if some one-dimensional coherence is maintained by exploiting ordering constraints [111, 73]. The former strategy is used in most

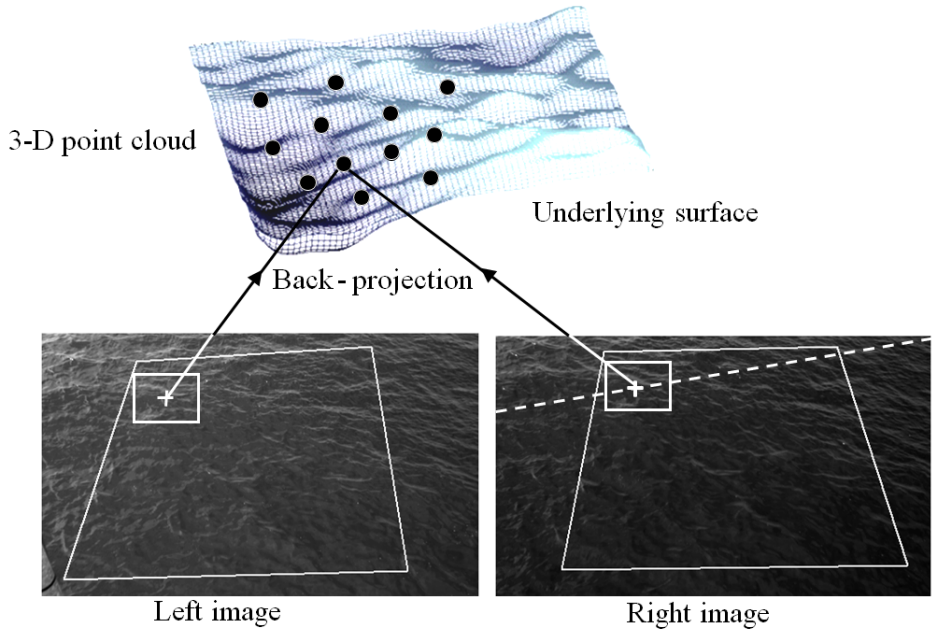


Figure 5: Classical stereo algorithms use feature matching along the epipolar line by optimizing a photometric score. Once a correspondence is established, the 3-D point is obtained by back-projection of the image points. Each 3-D point is obtained independently, resulting in a collection of scattered 3-D points (a *point cloud*).

recent systems applied to ocean engineering [4, 107]. This approach has the advantages of being simple to implement and fast to execute. However, it also has some major disadvantages that motivated the research for improved 3-D reconstruction methods [19, 110, 46]. These include:

- Each 3-D point is obtained independently and therefore the reconstructed surface of an object (water in our application) is obtained as a set of scattered 3-D points, sometimes also known as a "point cloud" in the literature. Thus, the hypothesis of the continuity (coherence) of the surface is not taken into consideration in the reconstruction process.

- A surface must be fitted to the scattered collection of reconstructed 3-D points to recover the underlying continuous surface of the objects in the scene. The simplest method to do this is by joining the points with the aid of a polygonal mesh. However, more elaborated methods can be utilized to solve this problem. Considered on its own, this is a difficult problem.

- If the matching algorithm does not support sub-pixel accuracy for the location of points $\mathbf{x}_i$, the resulting disparities will be quantized, forcing the reconstructed 3-D points to lie only at a certain discrete collection of depth planes from the cameras. This does not match the continuum of depths in the real world.

Some of the disadvantages can be mitigated by post-processing clean-up steps. For example: median filtering or Poincaré maps can be applied to remove spurious mismatches; holes due to mismatched regions can be filled by surface fitting. More appealing than post-processing, however, is the design of more powerful stereo methods that do not require artificial post-processing.

A thorough evaluation of several two-camera reconstruction algorithms under the assumptions of pre-calibrated cameras and dense output representation is reported in [82], where they are referred to as dense two-frame correspondence algorithms. Multi-view stereo methods take advantage of the visual information of the scene captured from several viewpoints to improve the accuracy of the matching process, and, consequently, the accuracy of the reconstruction. An empirical evaluation of several recent multi-view stereo methods can be found in [85] (2006). Since then, research in the multi-view stereo area has been very active [92, 65, 27, 28, 2].

In practice, a two-camera setup will be used for all data collections used in this thesis. For this number of cameras, the previously mentioned feature-based, multi-view algorithms perform poorly reconstructing the sea surface, as it has been confirmed.

Remark. In every 3-D reconstruction methods, the quality and accuracy of the results depend on the calibration of the cameras. There are standard camera calibration procedures in the literature to characterize accurately the intrinsic and extrinsic parameters of the cameras [105, 12, 114, 59]. The Camera Calibration Toolbox [6] is a friendly and efficient implementation of calibration methods and is available as part of standard computer vision open-source software like the OpenCV library [45]. It is not the objective of this thesis to address the details of such methods.

When the temporal variation of the ocean surface is also taken into account, a correct placement and spatial calibration of the cameras is essential to have maximum overlapping

fields-of-view (FOVs) of the cameras over the desired region of the ocean to be monitored (see Figure 3). In addition, camera synchronization during data acquisition is essential to have a consistent set of measurements in time. The maximum permissible time interval synchronization error is typically 5ms, (see [13, 43, 42]). In the rest of this thesis it is assumed that cameras are properly placed, calibrated and synchronized.

2.3 A variational geometric method

In this section, a variational method is presented to address the reconstruction problem following a *top-down* approach, which is philosophically opposite to the *bottom-up* approach used in traditional stereo methods [59, 35].

Consider the surface of the ocean as a surface supporting a Lambertian radiance function, i.e., radiance is independent of the viewing direction. This is approximately true on a cloudy day (uniform sky illumination hypothesis). At a given point on the water surface, the same amount of light is received from all possible directions and reflected toward the observer, causing a visual blurring of the specularities of the water. Under these conditions, traditional stereo algorithms (section 2.2) may fail to provide a smooth reconstructed surface because these algorithms are designed for surfaces with strong texture (significant intensity variations between neighboring points), which is not the case of the radiance of the water surface. This failure is evidenced by "holes" in the reconstructed surface, which correspond to unmatched image regions [59, 4]. Variational methods [19, 3, 110, 46] do not suffer from this shortcoming because they use a 2-D coherent model (either a 2-D disparity map between images or an explicit surface in 3-D space) of the smooth surface to be estimated, rather than representing it as a collection of scattered 3-D points. Another possible advantage of the variational approach over classical stereo is that a back-projection model might not be needed if 3-D point triangulation is not explicitly performed [19, 110, 46]. In the latter case, only the knowledge of the forward projection model is necessary, i.e., how to project a 3-D point on a given image plane. On the other hand, if a disparity map is used to represent the surface in a two-camera setup [3], the back-projection model is not needed during the search for the solution of the correspondence problem, but it is needed

afterwards during the depth-recovery step.

2.3.1 Energy functional minimization

The pioneering work in object centered, multi-view stereo reconstruction by Faugeras and Keriven [19], applies variational principles to reformulate the 3-D reconstruction problem of an object's surface into a surface evolution problem. A smooth surface $S \subset \mathbb{R}^3$ representing the desired object is obtained in the context of active surfaces by deforming an initial surface S_0 via a PDE derived from the gradient descent flow of an energy functional designed for the stereo reconstruction problem. The energy being optimized is an image-to-image intensity matching criterion in the form of a weighted area:

$$E(S) = \int_S \Phi(\mathbf{X})\, \mathrm{d}A, \tag{4}$$

where $\mathrm{d}A$ denotes the Euclidean area element of the surface. Energy functionals of this form are geometric in nature since the area element is invariant to different parameterizations of the same surface. The positive function $\Phi(\mathbf{X})$, where $\mathbf{X} \in S$, accounts for the stereo matching criterion: it is small at good matching locations and large otherwise.

Once the energy has been specified, a surface S that minimizes (4), at least locally, can be found via a gradient descent evolution introducing an artificial time marching variable t:

$$\begin{cases} \frac{\partial S}{\partial t} \equiv S_t = V_N \mathbf{N} \\ S(t=0) = S_0. \end{cases} \tag{5}$$

The speed in the normal direction to the surface, V_N, is obtained from the Euler-Lagrange (EL) equations of the designed energy, which are the necessary optimality conditions for functionals in infinite-dimensional spaces (see appendix A).

It is clear that the energy depends on the shape of the water surface via $\Phi(\mathbf{X})$. Therefore, the active surface establishes an evolving correspondence between the pixels in both images, as illustrated in Figure 6. Hence, the best correspondence will be obtained by deforming a surface in 3-D according to a PDE instead of only performing an epipolar line search without an explicit 3-D model of the target surface being reconstructed (section 2.2). According to the design of the energy functional, as the evolving surface approaches the true surface

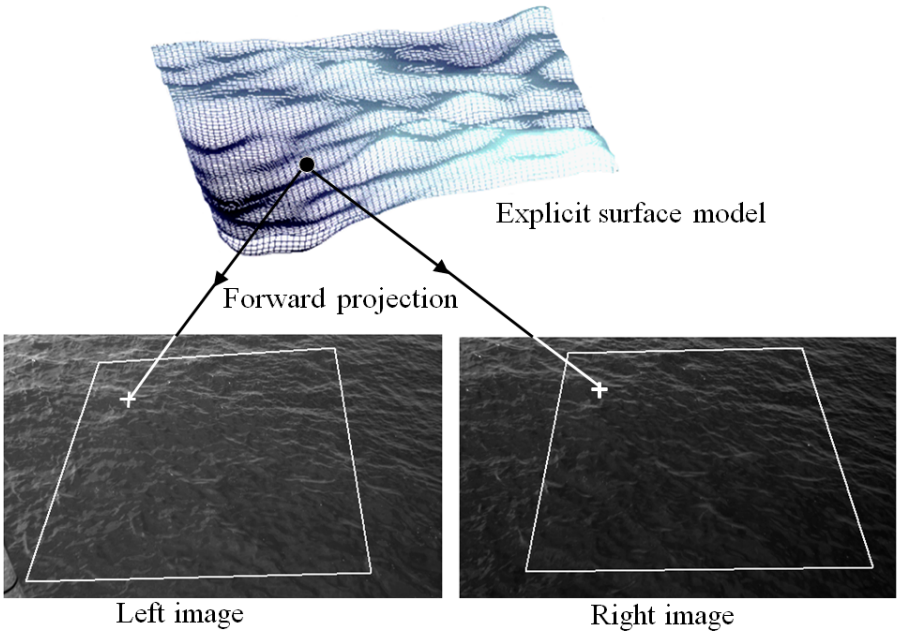


Figure 6: Variational stereo algorithms use an explicit model of the solution surface. The forward projection of surface points establishes correspondences between projected image points. The surface is deformed according to a gradient descent flow (PDE) that minimizes an energy designed to measure the photometric fit of a candidate surface to the observed images.

of the scene object, the velocity V_N decreases until finally, surface evolution halts when equation (5) reaches steady state.

A naive choice for $\Phi(\mathbf{X})$ is shown in equations (6)-(7), the 2-norm of the image errors between each pair of image intensities at the locations in which a surface point $\mathbf{X}$ projects, because this choice penalizes individual point mismatches in the image data and is therefore very sensitive to noise and local texture [19, 48].

$$\Phi(\mathbf{X}) = \frac{1}{n}\sum_{i \neq j} \Phi_{ij}(\mathbf{X}) \tag{6}$$

$$\Phi_{ij}(\mathbf{X}) = (I_i(\mathbf{x}) - I_j(\mathbf{x}'))^2 \tag{7}$$

A better choice for $\Phi(\mathbf{X})$ is a local matching criterion that does not involve comparisons among individual points, but instead involves comparisons of small neighborhoods around these points. In particular, in [19], a normalized cross-correlation measure between image intensities is used. An improved version of the matching criterion is described in [46], where image intensities are obtained by projecting the same surface patch onto both image

planes of the cameras, instead of re-projecting a surface patch defined in an arbitrary camera. In [29] it is shown how this variational approach and correlation-based choice for $\Phi(\mathbf{X})$ can be applied to the estimation of the water surface from stereo images. The details of the equations of this method are also explained therein, as well as their numerical implementation and their statistical and experimental validation.

2.3.2 Numerical solution. Level set methods

The numerical implementation of equation (5) proposed in [19] relies on the Osher-Sethian level set framework [69], where the surface $S(t)$ is implicitly represented as the zero level set of a higher dimensional function $\psi(X, Y, Z, t) : \mathbb{R}^4 \to \mathbb{R}$, called the level set function:

$$S(t) = \{(X, Y, Z) \in \mathbb{R}^3 \mid \psi(X, Y, Z, t) = 0\}. \tag{8}$$

The unit normal to the surface in equation (5) is, in terms of the level set function,

$$\mathbf{N} = -\frac{\nabla\psi}{\|\nabla\psi\|}, \tag{9}$$

where $\nabla\psi$ stands for the gradient with respect to the spatial arguments of ψ. Now, the evolution of the surface S described by (5) is achieved by evolving the level set function ψ according to the PDE

$$\psi_t = V_N \|\nabla\psi\|. \tag{10}$$

This PDE is usually initialized with a level set function ψ_0 that is the signed distance function to the surface. The level set function ψ is discretized in a 3-D grid enclosing the volume of interest in the world where the water surface is known to be. A numerical approximation of both sides of equation (10) is carried out using finite difference formulas. After the evolution, the surface S is extracted using the marching cubes algorithm [57], a popular algorithm to extract isosurfaces of volumetric data.

This framework has the advantage of easily handling topological changes during the evolution of the surface. Unfortunately, it also has a common disadvantage shared by other voxel-based methods [86]: the so-called dimensionality curse, i.e., the need to represent and evolve an entity by means of another one with a higher dimensionality.

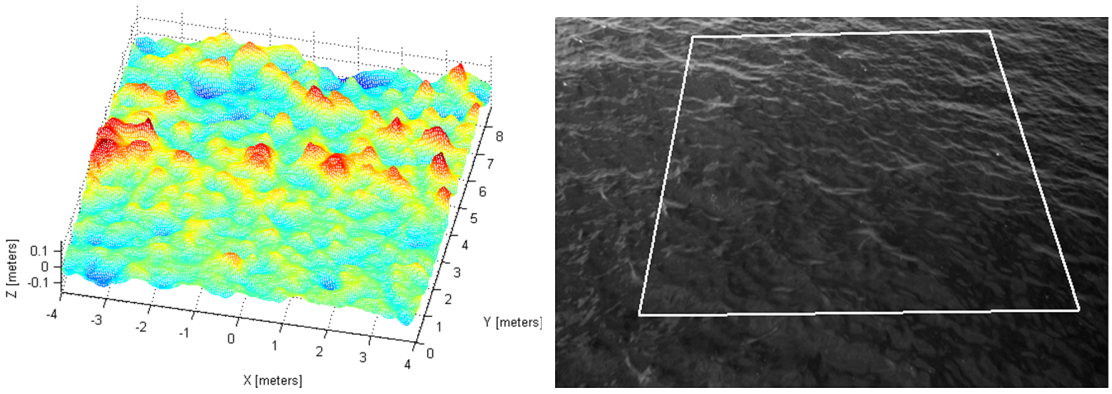


Figure 7: Left: water surface in a region of interest of $8m \times 8.7m$, reconstructed according to the variational geometric method [29]. The height of the waves is in the range $\pm 0.2m$. Right: region of interest projected on one of the images of the stereo pair.

The computational burden imposed by the dimensionality curse is partly reduced by evolving only a narrow band of the level set function around the surface [10, 11, 1]. In [70], a pyramidal technique is also suggested to ameliorate the computational burden of the level set evolution: solve the problem at a coarse resolution and interpolate the rough solution as an initial condition on a finer grid. In the level set framework, these are the two main accelerating techniques, which are efficient ways of implementing the steepest descent equation (10) derived from the EL equations of the proposed energy functional. Abandoning the idea of the level set representation of a surface, faster solvers will be proposed in section 3.4.2 for the particular case of surfaces representable as graphs or elevation maps.

In early exploration stages of my research, the level set surface evolution method previously described was tested on water stereo image data. The base code for the algorithms was provided by Dr. Jin [46] and the image data was provided by Dr. Benetazzo [4]. Some experiments were carried out and promising results have been reported [29, 20, 23]. The images of waves, on water depth of 8 meters, were cropped to 504×336 pixels to focus in the region close to the cameras. The distance between the optical centers of the cameras is $3.04m$ and the reconstructed surface occupies a rectangle of approximately $8 \times 8.7\, m^2$. Figure 7 shows the reconstructed surface of the water and the corresponding domain in one of the images of the stereo pair.

2.4 Dense disparity map estimation via variational methods

An alternative variational stereo method that lies in between classical image-based stereo methods and object-based stereo methods like [19] was presented by Alvarez et al. [3]. Their method follows the two-step scheme of classical (bottom-up) stereo methods described in section 2.2: they focus on solving the stereo matching problem between two images and they use standard triangulation to obtain the surface of the reconstructed object. However, they solve for the correspondence problem by enforcing 2-D coherence among the matched points. Each point (or epipolar line) is not estimated independently, but takes into account information from neighboring points (respectively, epipolar lines). The 2-D disparity map or disparity function establishes the correspondence between points of both images. Alvarez et al. use a continuous model of the disparity map and design an energy functional whose minimizer is the disparity map that solves the reconstruction problem (in a Lambertian scene).

The disparity map is a vector field with two components (the displacements in both horizontal and vertical directions of the image), however taking advantage of the epipolar constraint, the disparity can be parameterized by a single displacement: the signed distance along the epipolar line (denoted by λ in [3]), also called as the tangential component of the disparity. That is, given a point $\mathbf{x}_1$ and the fundamental matrix $\mathtt{F}$, the corresponding point $\mathbf{x}_2$ lies on the epipolar line $\mathbf{l}_2 = \mathtt{F}\bar{\mathbf{x}}_1$. The disparity $\mathbf{d} = \mathbf{x}_2 - \mathbf{x}_1 = \gamma\mathbf{N} + \lambda\mathbf{T}$ (in Euclidean coordinates) can be decomposed as the sum of two vectors in the normal ($\mathbf{N}$) and tangential ($\mathbf{T}$) directions to the epipolar line: the vector perpendicular to $\mathbf{l}_2$ starting at $\mathbf{x}_1$ and ending in $\mathbf{l}_2$, and the vector along the epipolar line ending in $\mathbf{x}_2$ (tangential component). This convenient parameterization allows the energy to be designed so that it depends on a single 2-D function, $E(\lambda)$ instead of two $E(\mathbf{d})$. The energy $E = E_{\text{data}} + E_{\text{smooth}}$, consists of: a data fidelity term E_{data} that measures the image-to-image discrepancy caused by a candidate disparity map, and a smoothness regularizer E_{smooth} that enforces coherence (continuity) of the disparity map. The PDE given by the Euler-Lagrange equations of the energy functional provide the necessary optimality conditions that the disparity map must satisfy

to be a minimizer of the energy. Such a non-linear PDE is solved numerically using time-stepping methods in a scale-space framework [108]. To solve for the disparity function, the two cameras need only be weakly calibrated, that is, only the knowledge of the fundamental matrix between both images is required. In other words, no Euclidean 3-D information is needed to establish correspondences; such information is only required in the triangulation phase to obtain the surface in a Euclidean coordinate system.

The main advantage of this method over other classical stereo methods is that, due to the 2-D coherence imposed on the disparity map, the densely reconstructed surface is smoother than that obtained by other means. Such coherence derives from the fact that variational principles allow the problem to be easily modeled in a continuous setup rather than a discrete one. Other benefits of the variational disparity method in [3] are its superior accuracy over purely cross-correlation matching, and the lack of interior unmatched regions ("holes") across both images.

Nevertheless, the variational disparity method presents some disadvantages. Apart from the computational burden imposed by the need to solve a non-linear PDE, the other main disadvantage is that depth of the scene is not taken into account since the cameras need only be weakly calibrated; to some extent we may say that the algorithm is purely image-based. This has terrible consequences in scenes like the ones in the ocean wave reconstruction problem. Here, images present a large depth variation since the image planes of the cameras and the approximately mean plane through the surface of the ocean are far from being in a parallel configuration. From the point of view of the regularizer of the energy, displacements of equal values have the same penalty regardless of where in the images they occur. Depth of the scene is not taken into account to favor (or penalize) large displacements for points that are close to (respectively, far from) the cameras, which would be a desirable situation. Another disadvantage is that once the stereo matching problem is solved by the dense disparity map, the 3-D surface still needs to be constructed. This can be expensive since triangulation needs to be performed on each point of the discretized disparity map. Moreover, the underlying continuous surface has to be fitted to the collection of scattered (but coherent) 3-D points. The method presented in this thesis (chapter 3) is designed to

take into consideration the depth of the objects in the scene, does not require triangulation, and builds an explicit model of the reconstructed surface, thus not requiring surface fitting to a set of points.

2.5 Conclusion

In this chapter we have elaborated on the two categories of pre-calibrated reconstruction algorithms mentioned in chapter 2 that are of interest to our ocean engineering application: image-based vs. object-based methods. The common multi-view image geometry and notation has been established in section 2.1. It will be used throughout this thesis. The differences between image-based and object-based stereo methods have been analyzed. In general, these methods follow two philosophically opposite approaches: bottom-up vs. top-down, respectively. Variational stereo is more powerful, flexible, and rigorous, albeit computationally expensive, than earlier traditional feature-based stereo methods based upon epipolar line search, matching, and back-projection to find the reconstructed surface of objects in the scene. However, the latter is easier to implement, whereas the former requires more mathematical background, especially in the context of partial differential equations and additional numerical techniques that are still the subject of active research.

CHAPTER III

THE VARIATIONAL GEOMETRIC METHOD. GRAPH SURFACE REPRESENTATION

The mathematical background of the variational framework in section 2.3.1 allows for the design of many energy functionals of the form (4). In this section, we present a different approach to the reconstruction problem by exploiting the hypothesis that the surface of the water can be represented in the form of a graph or elevation map:

$$Z = Z(X, Y), \tag{11}$$

where Z is the *height* of the surface with respect to a domain plane that is parameterized by coordinates X and Y.

Indeed, slow varying, non-breaking ocean waves admit this simple representation. As a natural extension of previous methods, energy functionals can be tailored to exploit the benefits of this valuable representation. Thus, the surface can still be obtained as the minimizer of a suitable energy functional but now with a different geometrical representation of the solution.

3.1 Advantages and disadvantages of the graph representation

The graph representation of the water surface presents some clear advantages over the level set representation of section 2.3:

- Discretization and dimensionality. Memory-wise, the surface is no longer represented as the zero level set of an auxiliary higher dimensional function ψ. Instead, it is represented as a collection of height values Z over a 2-D grid defined on the $X - Y$ plane. Only the height values need to be stored because the X and Y coordinates of the grid points will remain fixed during the evolution. The level set approach is more general because it can handle surfaces not representable as graphs. There is no added computational burden due to the evolution of the higher dimensional function

$\psi : \mathbb{R}^4 \to \mathbb{R}$ since now the surface will be evolved directly through the height function $Z(X, Y, t) : \mathbb{R}^3 \to \mathbb{R}$.

- The latter also implies that for the same amount of physical memory, higher spatial resolution (finer details) can be achieved in the graph representation than when the level set is used.

- Functions defined on the surface, e.g., the radiance f, admit a natural parameterization: $f(X, Y)$. Given a point in the $X - Y$ plane, it is straightforward to know what are the coordinates of the corresponding surface point $\mathbf{X} = (X, Y, Z(X, Y))^\top$ and the radiance at that point $f(X, Y)$. The $X - Y$ plane becomes the natural common domain to parameterize the geometrical and photometric properties of surfaces. This simple identification does not exist in the level set approach [110], where the shape and radiance of the surface are evolved in two 3-D grids since it is the only sensible way to maintain an easy mapping between both functions.

- The graph representation allows for fast numerical solvers besides gradient descent, like Fast Poisson Solvers [40, 32, 80], Cyclic Reduction [8, 31], Multigrid Methods [63, 7, 33], Finite-Element Methods (FEM) [96], etc. In the level set framework, the range of solvers is not as diverse.

However, there are also some minor disadvantages:

- A world coordinate properly oriented must be defined before running the evolution equation that drives the active surface. In particular, the domain plane $Z = 0$ (see Figure 8) must be determined to allow the water surface to be represented as a graph with respect to this plane. This is not trivial a priori and might pose a problem if only the information from the stereo images is used. This condition may not be so if external gravity sensors provide this information. It is also possible to choose an initial estimate for the plane and then update it with some feedback from the statistics of the reconstructed waves.

- Surface evolution is now constrained to be in the form of a graph since the driving

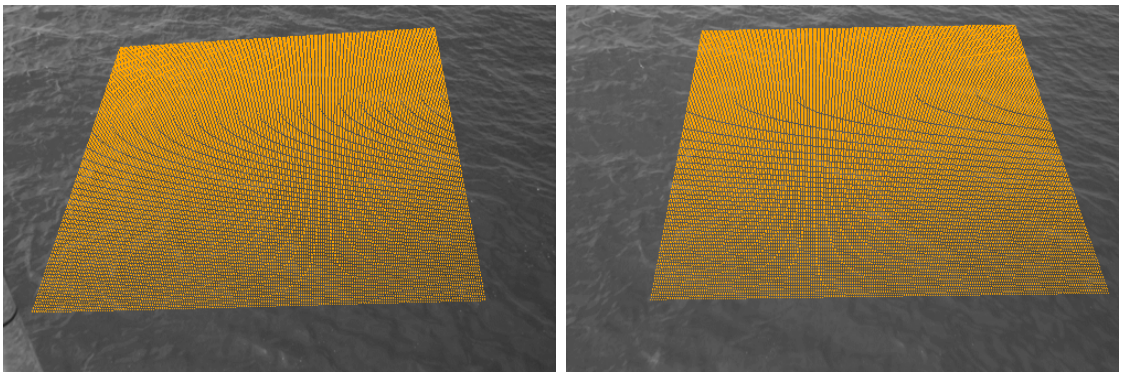


Figure 8: Projection of the initial model surface on a stereo image. The initial surface is a rectangle in the plane $Z = 0$, sampled in 210×210 grid points.

equation will be the resulting flow for the height function Z, which is derived from the EL equations. This may not be the same as the driving function dictated by an unconstrained surface evolution, as shown in appendix C. As a result, more iterations may be required to reach a local minimum.

The problem is mathematically stated in the following section. The desired surface is given by the solution of a variational optimization problem in the *height* of the surface.

3.2 Proposed energy functional

The variational stereo method in this thesis is inspired by a combination of both object-based stereo methods of Faugeras and Keriven [19] (FK), and Yezzi and Soatto [110] (Stereoscopic Segmentation - SS). Both are variational *top-down* approaches to multi-view stereo reconstruction / segmentation.

A key ingredient that makes our variational stereo method more robust against image noise than other stereo methods is the adoption of a *generative image model* for scenes with smooth appearance, as presented in SS. Both, the shape of the surface (geometry) and the color or grayscale intensity sitting on the surface (radiance) completely determine the intensity at a given image point via the perspective camera model and the calibration parameters. Images generated from these key components are called *modeled images*. Thus, the photometric error in SS is not computed by comparing image intensities directly, but indirectly via a (smooth) radiance model. SS is more robust to image noise than the

FK method because the latter relies on image-to-image intensity comparisons rather than image-to-model (i.e., image-to-radiance model) comparisons: in the FK method there is no notion of a radiance function defined on the surface of the object. The latter is equivalent to having a non-regularized radiance function that is a point-wise average of the image intensities according to the configuration of the cameras encoded in the epipolar geometry. This implicit radiance model is very sensitive to noise, as pointed out in section 2.3.1.

Another key ingredient of our method borrowed from SS is that the *photometric mismatch* between the images and a candidate surface is *measured in the images*, not on the surface (as in FK). By designing the energy functional in a suitable way, all image derivatives vanish and only derivatives of the radiance model appear [93], which can be controlled via a regularizer. Thus, noise amplification due to derivatives can be controlled; a property that the FK method lacks. Moreover, although the photometric criterion in SS is a point-wise comparison of image-to-model intensities, by using a regularizer on the radiance function one can also accomplish global neighborhood effects without having to choose arbitrary patches sizes and shapes (as in the FK method).

On the other hand, SS requires the notion of a *background* to segment the object(s) in the scene that best explains the given images. This is a limitation for our application of ocean waves, since the projection of the water surface may fill the entire domain of the images. Besides, we are not interested in solving the segmentation problem. The FK method lacks these limitations and therefore, it is a property that is borrowed in our variational stereo method.

A generative model of the images consisting of the joint estimation of the shape of the surface S and the radiance function on the surface f has been investigated. Consider the 3-D reconstruction problem from a pair of still images, i.e., do not consider temporal variation. Let the energy functional be the weighted sum of a data fidelity term E_{data} and two regularizing terms:

$$E(S, f) = E_{\text{data}}(S, f) + \alpha E_{\text{geom}}(S) + \beta E_{\text{rad}}(f), \qquad \alpha, \beta > 0. \tag{12}$$

The geometry smoothing term E_{geom} and the radiance smoothing term E_{rad} state that the

desired minimizer of the energy functional consists of smooth functions. The choice of these terms and the weights α, β, control the smoothness of the energy minimizer.

The data fidelity term is designed to measure the photo-consistency of the model. This term is the sum, over all images, of the squared differences between observed intensities I_i and those given by the radiance model f:

$$E_{\text{data}} = \sum_{i=1}^{N_c} E_i, \tag{13}$$

where the i-th image-to-model comparison is

$$E_i = \int_{\Omega_i} \phi_i \, \mathrm{d}\mathbf{x}_i, \tag{14}$$

and the photometric matching criterion is

$$\phi_i = \tfrac{1}{2}\big(I_i(\mathbf{x}_i) - f(\mathbf{x}_i)\big)^2. \tag{15}$$

The region of the i-th image where the scene is projected is denoted by Ω_i. Assuming that the surface of the scene (water) can be represented in the form of a graph $Z = Z(u, v)$, a point on the surface has coordinates

$$\mathbf{X}(u, v) = \big(u, v, Z(u, v)\big)^\top. \tag{16}$$

The chain of transformations to obtain the intensity $I_i(\mathbf{x}_i)$ given a point $\mathbf{u} = (u, v)^\top$ in the parameter space of the surface is

$$\mathbf{u} \mapsto \mathbf{X}(\mathbf{u}) \mapsto \tilde{\mathbf{X}}_i = \mathtt{M}^i\mathbf{X} + \mathbf{p}_4^i \mapsto \mathbf{x}_i \mapsto I_i(\mathbf{x}_i), \tag{17}$$

where $\mathbf{X}(\mathbf{u}) \equiv S(\mathbf{u})$ are the world coordinates of a surface point, $\tilde{\mathbf{X}}^i = (\tilde{X}_i, \tilde{Y}_i, \tilde{Z}_i)^\top$ are related to the coordinates of the surface point $\mathbf{X}$ in the i-th camera frame, and $\mathbf{x}_i = (x_i, y_i)^\top = (\tilde{X}_i/\tilde{Z}_i, \tilde{Y}_i/\tilde{Z}_i)^\top$ are the coordinates of the projection of $\mathbf{X}$ in the i-th image plane. The projection matrix of the camera corresponding to the i-th image is, in world coordinates,

$$\mathtt{P}^i = [\mathtt{M}^i \mid \mathbf{p}_4^i], \tag{18}$$

where $\mathtt{M}^i = \mathtt{K}^i\mathtt{R}^i$ comprises the intrinsic parameters ($\mathtt{K}^i$) and the orientation of the camera (3-D rotation $\mathtt{R}^i$), and $\mathbf{p}_4^i = \mathtt{K}^i\mathbf{t}^i$ contains the camera's location (translation). The determinant of $\mathtt{M}^i$ is denoted by $|\mathtt{M}^i| = \det(\mathtt{M}^i)$. Because $\det(\mathtt{R}^i) = 1$, it follows that $|\mathtt{M}^i| = |\mathtt{K}^i|$.

The radiance model f is specified by a function $\hat{f}$ defined on the surface S. Moreover, we consider its extension to the whole embedding space $\hat{f} : \mathbb{R}^3 \to \mathbb{R}$. There are many possible ways to define this extension; we will consider one that simplifies the equations of the model. Then, f in (15) is naturally defined by $f(\mathbf{x}_i) = \hat{f}(\pi_i^{-1}(\mathbf{X}))$, where π_i^{-1} denotes the back-projection operation from a point in the i-th image to the closest surface point with respect to the camera. With a slight abuse of notation, let us use f to denote the parameterized radiance $f(\mathbf{u})$, understanding that $f(\mathbf{x}_i)$ in (15) reads the back-projected value in $\hat{f}(\mathbf{X}(\mathbf{u})) = f(\mathbf{u})$.

Motivated by the common parameterizing domain of the shape Z and radiance f of the surface, and seeking to obtain the simplest diffusive terms in the PDEs that derive from the necessary optimality conditions of the energy (12), let the regularizers (also known as *regularity penalties*) be

$$E_{\text{geom}} = \int_U \tfrac{1}{2} \|\nabla Z(\mathbf{u})\|^2 \, \mathrm{d}\mathbf{u}, \tag{19}$$

$$E_{\text{rad}} = \int_U \tfrac{1}{2} \|\nabla f(\mathbf{u})\|^2 \, \mathrm{d}\mathbf{u}. \tag{20}$$

Now that all terms in (12) have been specified, some transformations are carried out to express the integrals over a more suitable domain. Integrals in (19) and (20) are already in a convenient domain, the parameter space. The data fidelity term (13) can be expressed as an integral over the parameter space by means of a change of variables. Let the Jacobian of the change of variables be (see (181), (185) in appendix D.1)

$$\mathtt{J}_i = \left| \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}} \right| = -|\mathtt{M}^i| \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot (\mathbf{X}_u \times \mathbf{X}_v). \tag{21}$$

Then, the data fidelity energy (14) becomes

$$E_i = \int_{\Omega_i} \phi_i \, \mathrm{d}\mathbf{x}_i = \int_U \phi_i \mathtt{J}_i \, \mathrm{d}\mathbf{u}, \tag{22}$$

where the last integral is over U: the part of the parameter space whose surface projects on Ω_i in the i-th image.

Furthermore, the data fidelity term can be expressed as a surface integral, according to the relationship between area measures (184) (see appendix D.2), with

$$E_i = -\int_S \phi_i(\mathbf{X}, f) |\mathtt{M}^i| \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot \mathbf{N} \, \mathrm{d}A. \tag{23}$$

A visibility term (in the form of a characteristic function) that states what part of S should be integrated according to what part of S is visible from the i-th camera must also be included in the integrand of (23), but it has been omitted for simplicity.

After collecting terms (19), (20), and (22), energy (12) is in the form

$$E(\mathbf{X}, f) = \int_U L(\mathbf{X}, \mathbf{X}_u, \mathbf{X}_v, f, f_u, f_v, u, v)\, \mathrm{d}\mathbf{u}, \tag{24}$$

where subscripts indicate the derivative with respect to that variable, and the integrand is the so-called *Lagrangian*:

$$L = L_{\text{data}} + \alpha L_{\text{geom}} + \beta L_{\text{rad}}, \tag{25}$$

with

$$L_{\text{data}} = \sum_{i=1}^{N_c} L_i, \tag{26}$$

$$L_i = \phi_i \mathrm{J}_i, \tag{27}$$

$$L_{\text{geom}} = \tfrac{1}{2}\|\nabla Z(\mathbf{u})\|^2, \tag{28}$$

$$L_{\text{rad}} = \tfrac{1}{2}\|\nabla f(\mathbf{u})\|^2. \tag{29}$$

3.3 Energy minimization. Optimality condition

The energy (24) depends on two functions: the shape of the surface $\mathbf{X}$ and the radiance function on the surface f. To find a minimizer of such a functional, we derive the necessary optimality condition by setting to zero the first variation of the functional. Before that, let us enforce the constraint on the representation of the surface as a graph (16). Fixing the parameterization $\mathbf{u}$, the energy functional (24) becomes a function of the height $Z(\mathbf{u})$:

$$E(Z, f) = \int_U L(Z, Z_u, Z_v, f, f_u, f_v, u, v)\, \mathrm{d}\mathbf{u}. \tag{30}$$

The first variation (Gâteaux derivative) of (30) is

$$D_{(h,w)}E(Z, f) \equiv \frac{\delta E}{\delta(Z, f)}(h, w) = \left.\frac{\mathrm{d}E\big((Z, f) + \epsilon(h, w)\big)}{\mathrm{d}\epsilon}\right|_{\epsilon=0}, \tag{31}$$

where h and w are arbitrary (but smooth) perturbations for Z and f, respectively.

Using the chain rule and integration by parts in (31), it follows that the first variation (32) has two terms: one in the interior of the integration region U in the parameter

space and one boundary term (on ∂U), where $\boldsymbol{\nu} = (\nu^u, \nu^v)^\top$ represents the outward unit normal to U:

$$\begin{aligned}\frac{\delta E}{\delta(Z,f)} = &\int_U \Big(\big(L_Z - (L_{Z_u})_u - (L_{Z_v})_v\big) \cdot h + \big(L_f - (L_{f_u})_u - (L_{f_v})_v\big) w \Big)\, \mathrm{d}\mathbf{u} \\ &+ \int_{\partial U} \Big(\big(L_{Z_u}\nu^u + L_{Z_v}\nu^v\big) \cdot h + \big(L_{f_u}\nu^u + L_{f_v}\nu^v\big) w \Big)\, \mathrm{d}\sigma.\end{aligned} \tag{32}$$

Setting the first variation to zero for all possible perturbations (h, w) yields a coupled system of PDEs along with natural boundary conditions:

$$L_Z - (L_{Z_u})_u - (L_{Z_v})_v = 0 \qquad \text{in } U, \tag{33}$$

$$L_{Z_u}\nu^u + L_{Z_v}\nu^v = 0 \qquad \text{on } \partial U, \tag{34}$$

$$L_f - (L_{f_u})_u - (L_{f_v})_v = 0 \qquad \text{in } U, \tag{35}$$

$$L_{f_u}\nu^u + L_{f_v}\nu^v = 0 \qquad \text{on } \partial U. \tag{36}$$

After some calculations (see appendix D) equations (33), (34), (35), and (36) become

$$g(Z,f) - \alpha\Delta Z = 0 \qquad \text{in } U, \tag{37}$$

$$b(Z,f) + \alpha\frac{\partial Z}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \partial U, \tag{38}$$

$$-\sum_{i=1}^{N_c}(I_i - f)\mathtt{J}_i(Z) - \beta\Delta f = 0 \qquad \text{in } U, \tag{39}$$

$$\beta\frac{\partial f}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \partial U, \tag{40}$$

where the non-linear terms due to the data fidelity energy are

$$g(Z,f) = \nabla f \cdot \sum_{i=1}^{N_c} |\mathtt{M}^i| \tilde{Z}_i^{-3}(I_i - f)(u - C_i^1, v - C_i^2), \tag{41}$$

$$b(Z,f) = \sum_{i=1}^{N_c} \phi_i |\mathtt{M}^i| \tilde{Z}_i^{-3}\big((u - C_i^1)\nu^u + (v - C_i^2)\nu^v\big). \tag{42}$$

The Laplacians ΔZ and Δf arise from the regularizing terms (19) and (20), respectively, and $\partial * / \partial \boldsymbol{\nu}$ is the usual notation for the directional derivative along $\boldsymbol{\nu}$, the normal to the integration domain U in the parameter space.

The calculations are involved, but a simple classification of both diffusive PDEs can be done as follows. On the one hand, for a fixed surface, (39) and (40) form a linear elliptic PDE (of the inhomogeneous Helmholtz type) with homogeneous Neumann boundary conditions.

On the other hand, for a fixed radiance, (37) and (38) lead to a non-linear elliptic equation in the height Z with nonstandard boundary conditions.

Observe that if there was no regularizing term on the radiance ($\beta = 0$), equation (39) would be linear in f, and the solution would be a weighted average of the intensities at the image projections of the surface (188). The motivation to include the term E_{rad} in (12) and, therefore, to use a diffusive PDE in the radiance function supported by the surface shape is to increase the basin of attraction for the shape minimizer, thus avoiding getting trapped in local minima of the energy.

Also observe the remarkable fact that no derivatives of the image data appear in the final expression of the EL equations. Moreover, the derivatives have been transferred to the radiance model, whose smoothness is controlled by the regularizer E_{rad}. This desirable property makes the algorithm less sensitive to image noise when compared to other variational approaches for stereo 3-D reconstruction. This property is shared by the standard Mumford-Shah [66] formulation for direct image segmentation. In our case, this property arises from two facts: (i) our model includes a radiance function f, and (ii) the stereo discrepancy is measured in the image domain rather than on the surface [93]. Moreover, it is a purely geometric result, thus independent of the choice of the shape of the error function ϕ_i in case that a non-quadratic penalty is chosen, as it will be discussed in section 5.3.

The following discussion reveals the importance of the proper choice of the weights in the composite energy functional to achieve both a visually and a physically meaningful solution. Roughly, α and β control the amount of high frequencies in Z and f, respectively, that are allowed in the reconstructed surface. As $\beta \to 0$, the radiance approaches the crisp weighted linear solution (188). As β increases, the radiance becomes smoother, according to the increasing diffusion in the PDE (39). However, a big value of β imposes too much smoothness on the radiance, so that even a small α (a non-smooth surface) is not able to overcome the constraint and allow the decrease of the data fidelity energy. A value $\alpha \neq 0$ generates the spatial coupling or coherence of the reconstructed surface shape, according to the diffusion process in (37). Since α determines the amount of smoothness of the reconstructed surface shape, it affects the statistics of the latter. This is better observed

in the omni-directional spectrum of the height function (section 6.2.4.1). A big α (smooth surface) causes a steep decay of the omni-directional spectrum in the high frequency region. As $\alpha \to 0$, high frequency components rise, and the slope of the tail becomes milder. In practice, α is empirically determined to yield the correct physical slope in the spectrum of the reconstructed surface according to Zakharov's theory [112].

Only in a limited number of simple cases it is possible to find an analytic and simple solution to the EL equations. In most image processing applications it is not possible to solve such equations directly. Instead, a common approach to solve difficult EL equations, such as the ones presented in (37)-(40), is to add an artificial time marching variable t dependency in the unknown functions (height, radiance), and set up a gradient descent flow that will drive the evolution of a candidate initial solution so that the energy (30) will decrease in time. Thus the solution of the elliptic PDEs (EL equations) is obtained as the steady state of the gradient descent equations. This is the context of the so-called active surfaces. The gradient descent equations are:

$$Z_t = \alpha \Delta Z - g(Z, f), \tag{43}$$

$$f_t = \beta \Delta f - \sum_{i=1}^{N_c} \mathtt{J}_i(Z)\, f + \sum_{i=1}^{N_c} I_i \mathtt{J}_i(Z). \tag{44}$$

To simplify the equations, we approximate the boundary condition (38) by a simpler, homogeneous Neumann boundary condition. This can be interpreted as if the data fidelity term vanished close to the boundary, and it is a reasonable assumption since the major contribution to the energy is given by the terms in the interior of the discretized domain, not at the boundary.

3.4 Numerical solution

An iterative method is used to find the minimum of energy (12) via the solution of the coupled system of equations that arise from the the necessary optimality conditions of the energy. Observe that, for a fixed height, equation (39) is a linear PDE in the radiance, which is much easier to solve than the non-linear PDE in the height (37) for a fixed radiance. By exploiting this asymmetry one may devise a minimization strategy consisting of a nested iterative scheme: an outer loop performing a gradient descent in the height (43), and an

inner loop implementing a direct optimization for the radiance. This is the method proposed in [49] to solve a coupled system of PDEs similar to the one proposed in this thesis. Starting from an initial approximate solution, there are two phases within each iteration: (1) compute the optimal radiance for a fixed shape, and (2) evolve the shape, leaving the radiance fixed. This decoupling strategy is a type of *operator splitting* [76] and it allows us to consider each equation, (37) and (39), as being in a single unknown (Z or f) between iterations.

The proposed PDEs are discretized on a rectangular 2-D grid in the parameter space and then solved numerically using finite-difference methods (FDM) [99]. To be consistent with the notation in [7], let us use the more familiar symbols (x, y) instead of (u, v) to refer to points in the parameter space. The computational grid consists of uniformly spaced points $(x_i, x_j) = (i\Delta x, j\Delta x)$, where Δx and Δy are the grid spacings in each dimension. We further assume equidistant grid steps $h = \Delta x = \Delta y$ to simplify the analysis. The standard notation of integer indices i, j and n in

$$q_{i,j}^{n} = q(i\Delta x, j\Delta y, n\Delta t), \quad \text{where} \quad q = \{Z, f\}, \tag{45}$$

is used to simplify the notation of the finite-difference equations resulting from the discretization of the PDEs. Therefore, in the aforementioned nested minimization approach, each decoupled PDE problem is cast in discrete form into a linear or non-linear system of equations. Let us first discuss the solution of the linear PDE for the radiance function since it is simpler than the solution of the PDE in the height function.

3.4.1 Linear PDE in the radiance function

Consider the numerical solution of the PDE in the radiance (39),

$$\beta\Delta f - cf + K = 0,$$

where, $c = \sum_{i=1}^{N_c} \mathtt{J}_i$ and $K = \sum_{i=1}^{N_c} I_i \mathtt{J}_i$, using FDMs. Replacing the second order derivatives of the PDE by finite differences leads to the system of linear equations

$$A\mathbf{u} = \mathbf{f}, \tag{46}$$

where $\mathbf{u}$ is the vector of unknowns (f_{ij}), $\mathbf{f}$ is the vector corresponding to the independent term K, and matrix A is a banded matrix resulting from the discretization of the Laplacian

operator and the linear term cf. This linear system also includes the discretization of the boundary conditions (40) that accompany the PDE. For example, if $v_{i,j}$ is an approximation to the exact solution $f(x_i, y_j)$, $c_{ij} = c(x_i, y_j)$, $K_{ij} = K(x_i, y_j)$, and the 5-point stencil is used as an approximation of the Laplacian,

$$(\Delta f)_{i,j} \approx \frac{1}{h^2}(v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{i,j}), \tag{47}$$

the finite-difference equation for a point in the interior of the grid is

$$\frac{\beta}{h^2}(v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{i,j}) - c_{i,j}v_{i,j} + K_{ij} \approx 0.$$

Therefore, the original linear PDE problem (39)-(40), is cast in discrete form into a *linear system* of equations. Standard tools of numerical analysis can be used to approximately solve this linear system efficiently (without having to explicitly compute the inverse matrix). One such tool is the Jacobi (simultaneous displacement) method [80]. For an interior point, this produces an iteration scheme that may be written in component form as

$$v_{i,j}^{(k+1)} = (4\tilde{\beta} + c_{i,j})^{-1}\big(\tilde{\beta}(v_{i+1,j}^{(k)} + v_{i-1,j}^{(k)} + v_{i,j+1}^{(k)} + v_{i,j-1}^{(k)}) + K_{ij}\big),$$

where $\tilde{\beta} = \beta/h^2$ and $v_{i,j}^{(k)}$ is the approximate solution at the k-th iteration. The scheme is initialized by some guess, which in our problem is the solution of the equation for $\beta = 0$. Convergence of the Jacobi method depends on the spectral radius of the iteration matrix, which ultimately derives from A. Strict diagonal dominance of A is a sufficient condition for convergence. In our case, A is diagonally dominant because $|4\tilde{\beta} + c_{ij}| \geq 4\tilde{\beta}$. The inequality is strictly satisfied if $c_{ij} > 0$, which is related to the necessary visibility condition for a point on the surface. A desirable stopping criterion of the iterations carried out to solve (46) would measure the relative error in the solution vector at the k-th iteration: $\|\mathbf{v}^{(k)} - \mathbf{u}\| < \epsilon\|\mathbf{u}\|$. In practice, however, $\mathbf{u}$ is unknown and A can be used to transfer the stopping condition to the norm of the residual vector $\mathbf{r}^{(k)} = \mathbf{f} - A\mathbf{v}^{(k)} = A(\mathbf{u} - \mathbf{v}^{(k)})$: $\|\mathbf{r}^{(k)}\| \leq \epsilon\|\mathbf{f}\|$.

Other standard methods like weighted Jacobi, Gauss-Seidel (regular, symmetric or red-black versions), Successive Overrelaxation (SOR), etc. may be used to solve the above linear system. See [51, 80] and the references therein for more details of the aforementioned methods. In this thesis, the Jacobi and red-black Gauss-Seidel (successive displacement)

Table 1: Complexity of different solvers for the 2-D Poisson problem (N denotes the total number of unknowns, i.e., grid points).

Method	♯ operations in 2-D
Gaussian elimination (band version)	$O(N^2)$
Jacobi iteration	$O(N^2 \log \epsilon)$
Gauss-Seidel iteration	$O(N^2 \log \epsilon)$
Successive overrelaxation (SOR)	$O(N^{3/2} \log \epsilon)$
Conjugate gradient (CG)	$O(N^{3/2} \log \epsilon)$
Nested dissection	$O(N^{3/2})$
Alternating Direction Implicit (ADI)	$O(N \log N \log \epsilon)$
Fast Fourier Transform (FFT)	$O(N \log N)$
Total reduction (Fast Poisson Solver)	$O(N)$
Multigrid (FMG)	$O(N)$

methods have been implemented and used as relaxation steps within the framework of multigrid methods [63, 7, 33]. These methods usually have better convergence rates than time-stepping methods for solving the gradient descent equation (44). In addition, the latter require additional work to compute an extra parameter: the time step between iterations. Time-stepping methods are discussed in section 3.4.3.

Due to the particular structure of the radiance problem (linear PDEs discretized on a Cartesian grid and homogeneous Neumann boundary conditions), fast methods like Fast Poisson Solvers [40, 32, 76, 80] and Cyclic Reduction [8, 31, 76] are among the most efficient numerical algorithms to address the problem. More specifically, since the PDE is discretized on a Cartesian grid and the Neumann boundary conditions are homogeneous, the Discrete Cosine Transform (efficiently implemented based on the Fast Fourier Transform - FFT) is at the heart of the Fast Poisson Solver that can tackle the problem. However, it was decided not to implement these specialized solvers in this thesis since multigrid methods achieve similar or better performance and are of more general applicability. Table 1, extracted from [103], gives an overview of the complexity of different solvers that can be applied to our problem. Direct and iterative solvers are listed. The $\log \epsilon$ term accounts for the discretization accuracy; the accuracy (stopping criterion) of the iterative methods is assumed to be of the same order.

3.4.2 Non-linear PDE in the height function

Analogously to the argument in the previous section, it is possible to seek a numerical solution of the PDE and boundary conditions in the height function using FDMs. However, the output of the discretization step will not be a linear system of equations due to non-linear nature of the PDE (37). Instead, it will yield a non-linear system of equations that can be written in the general form

$$A(\mathbf{u}) = \mathbf{f}, \tag{48}$$

where $\mathbf{u}$ represents the unknown height function at the grid points, A is an (non-linear) operator defined by the Laplacian and the non-linear term of the PDE, and $\mathbf{f}$ represents a possibly non-zero right side. The non-linear operator may also include the effect of the right side so that (48) becomes $\tilde{A}(\mathbf{u}) \doteq A(\mathbf{u}) - \mathbf{f} = \mathbf{0}$. However, to follow the notation in [7] for multigrid methods, the system is presented as in (48).

The solution of the non-linear problem is a lot harder than the linear problem. Nevertheless, references [7, 63, 76, 103] are an excellent guidance and source of multiresolution methods to tackle problems of this type. In particular, the multigrid methods called the full approximation scheme (FAS) and the full multigrid FAS (FMG-FAS) have been implemented to iteratively solve the non-linear system of equations efficiently. From a practical point of view, the main property of multigrid methods is that the rate of convergence is independent of the grid/mesh size, $h = \Delta x$.

Unlike other methods, multigrid can achieve computing work and storage proportional to the number of unknowns N in the problem, $O(N)$, not just for special cases, but for a broad class of problems. As the number of unknowns increases, multigrid becomes more efficient. For large problems, multigrid methods are currently the preferred solvers for elliptic boundary value problems because they are among the most efficient numerical tools.

Although multigrid methods have broad scope and applicability [33], they can be regarded as very efficient ways to solve linear systems of equations, e.g., those that arise from the discretization of PDEs. Multigrid can be considered a type of preconditioning or a technique to accelerate convergence of stationary iterative solvers (called smoothers or

relaxation methods in the multigrid context). Preconditioning techniques [17] can be used to accelerate multigrid methods. However, if the multigrid algorithm is well designed and fits the problem, it will converge very quickly, relegating preconditioning acceleration to a second place and granting the extra effort required to include preconditioning a worthless attempt [109].

Another possible approach to solve for the height function is to use the same idea as in Newton's method [52, 17]: approximate the non-liner PDE by a linearized version, solve the linear problem (which we know from the previous section), update the current estimate of the solution and keep on iterating until, hopefully, it converges. Using a Taylor series expansion around the current height estimate $\bar{Z}$, we may linearize the non-linear term (41),

$$g(Z) \approx g(\bar{Z}) + \dot{g}(\bar{Z})(Z - \bar{Z}), \tag{49}$$

resulting in, from (37), the linearized PDE:

$$\alpha \Delta Z - g(\bar{Z}) - \dot{g}(\bar{Z})(Z - \bar{Z}) \approx 0,$$

that is,

$$\alpha \Delta Z - \dot{g}(\bar{Z}) Z = g(\bar{Z}) - \dot{g}(\bar{Z})\bar{Z}. \tag{50}$$

Incorporating the homogeneous Neumann boundary conditions and discretizing, the problem is cast in discrete form into the *linear system* of equations:

$$A'\mathbf{u} = \mathbf{f},$$

where $\mathbf{u}$ is the vector of unknowns (Z_{ij}), $\mathbf{f}$ is the vector corresponding to the right side of (50) and matrix A' is a banded matrix resulting from the discretization of the Laplacian operator and the linear term $\dot{g}(\bar{Z})Z$. This linear system of equations can also be obtained by swapping the linearization and discretization steps, i.e., by linearizing the system of equations (48).

Once again, multigrid methods can be used to efficiently solve the linearized problem. Details of numerical implementation of these methods can be found in [7, 76]. According to the smoothing philosophy behind relaxation procedures in multigrid methods, the

coefficients of the linearized PDE are fixed within each V or W-cycle [7], that is, the current estimate of the solution $\bar{Z}$ is only updated in-between cycles. For the same number of multigrid cycles, the linearized approach may not be as accurate as the non-linear approach, but it outperforms the latter speed-wise because the non-linear term requires the bulk of the computational work and it only needs to be computed at the finest grid level in the linearized approach. Conceptually and implementation-wise, there is another important difference between both approaches: in the linearized approach, one only needs to know how to discretize the PDE (i.e., how to compute the coefficients of the linear PDE from the non-linear PDE) at the finest level because the coarser grids can be regarded as auxiliary computational grids with no physical meaning. In contrast, in the non-linear approach, one needs to know how to discretize the PDE at all levels, which can be troublesome depending on the form of the non-linearity.

3.4.3 Time-stepping methods. Stability analysis

Time-stepping methods are frequently used to solve PDEs describing stationary phenomena such as those described by elliptic PDEs. Time-stepping is a temporally discretized view of gradient descent PDEs like (43) and (44). The solution of the elliptic PDE is obtained as the steady-state of the time-stepping PDE. Next, we summarize the results of the ℓ^2 stability analysis carried out in appendix F for the numerical solution of the PDEs previously presented.

The stability requirements for a generic linear PDE of the form

$$f_t = \beta \Delta f - cf + K$$

that is discretized using an *explicit scheme* with forward finite differences in time,

$$f_t \approx \frac{f_{i,j}^{n+1} - f_{i,j}^{n}}{\Delta t}, \tag{51}$$

and central differences in space (FTCS), are

$$c_{i,j} \geq 0 \;\; \forall i,j \qquad \text{and} \qquad \Delta t \leq \big(\frac{4\beta}{h^2} + \frac{\max c_{i,j}}{2}\big)^{-1}. \tag{52}$$

The second requirement imposes a constraint in the maximum time step Δt allowed in the update equation, a typical property of explicit schemes.

We may apply the previous result to state the stability condition for the gradient descent PDE in the radiance function. Identifying coefficients,

$$c_{i,j} = \sum_{k=1}^{N_c} \mathrm{J}_{i,j}^k, \quad \text{and} \quad K_{i,j} = \sum_{k=1}^{N_c} I_{i,j}^k \mathrm{J}_{i,j}^k. \tag{53}$$

Since the Jacobian of the change of variables is non-negative, its point-wise discretization is also non-negative and the first requirement for stability is satisfied ($c_{i,j} \geq 0 \quad \forall i, j$). The maximum time step allowed in the update equation is (52), with c_{ij} given by (53). Therefore, the Jacobian of the change of variables (due to the data fidelity term in the energy functional) and the weight of the regularizer (β) totally characterize the stability requirements for the gradient descent PDE in the radiance function.

By a linearization argument, it is also possible to apply the above result to state an approximate stability condition for the time-stepping PDE in the height function with a Laplacian diffusive term:

$$\dot{g}(\bar{Z}_{i,j}) \geq 0 \ \ \forall i, j \qquad \text{and} \qquad \Delta t \leq \big(\frac{4\alpha}{h^2} + \frac{\max|\dot{g}(\bar{Z}_{i,j})|}{2}\big)^{-1}. \tag{54}$$

Under the assumption that the linearized PDE is a good local approximation of the original, non-linear PDE, the previous stability result is also used in the non-linear PDE. Therefore, the derivative of the non-linear term (due to the data fidelity term in the energy functional) and the weight of the regularizer (α) specify the stability condition of the linearized PDE in the height function. The expression for the derivative of the non-linear term due to the data fidelity energy can be found in appendix F.3.

It is possible to use other finite-difference schemes besides FTCS to discretize the PDE(s). For example, a backward difference approximation for the time derivative would yield an *implicit* scheme, where updating the approximate solution requires solving a linear system at every iteration. This is computationally more expensive than the explicit scheme, but it is unconditionally stable in time (at least in the linear case), which allows for larger time step Δt (usually, of the order $O(h)$ instead of $O(h^2)$).

3.4.4 Nested multigrid iteration

The previous time-stepping method for the height is used as a relaxation procedure inside a multigrid method that approximately solves the corresponding EL equation. The PDE in the radiance function is solved using Jacobi or Gauss-Seidel iteration since it is faster than time-stepping. This is summarized in Figure 9. The scheme can be interpreted as a nested loop: in the outer loop, the surface shape is evolved and in the inner loop, the optimal radiance is computed for the given shape. Usually, the number of V-cycles $NV_Z = NV_f = \{1, 2\}$ are reasonable values for a single iteration. Variational optimization methods like the one in Figure 9 usually require hundreds of iterations to reach convergence. From a practical point of view, convergence can be determined by detecting stagnation of the monitored energy or residual of the height PDE. In addition, sensible values of the number of pre- and post-relaxation sweeps [7] at each level of the multigrid solver are $\{1, 2\}$.

If the surface is initialized by a zero plane, the execution of the *nested iteration* procedure in FMG-FAS prior to the scheme in Figure 9 significantly accelerates convergence (see [7]) because the estimates from coarser resolutions provide a good initialization of the unknowns at the next finer resolution. Most of the computational work in an iteration is spent on the evolution of the non-linear PDE, e.g., step (1) of Figure 9. A V-cycle to evolve the radiance is faster than a V-cycle to evolve the shape because: (*i*) the radiance PDE is linear, (*ii*) the calculation of the coefficients of the radiance PDE require less computational effort than the calculation of the non-linear term of the height PDE, (*iii*) the stationary iterative smoother (Jacobi or Gauss-Seidel) converges faster than the time-stepping method for the height PDE.

From a practical point of view, the method has been implemented in C++ language because it offers, at the developing stage, a good compromise between speed of execution and object oriented modularity of the code. At the developing stage, mathematical tools such as multigrid methods have been adopted and proven successful to accelerate the convergence rate of the PDE solvers instead of using other brute-force approaches such as hardware acceleration via Graphics Processing Unit (GPU) or via parallel processing. These may be explored in future work.

Given an initial guess of the surface shape $Z(u, v)$ and radiance $f(u, v)$ (and, optionally, the AGC coefficients of the radiance extended model described in section 3.6), iterate according to the following alternating scheme until convergence:

1. Perform NV_Z V-cycles of multigrid to evolve the surface shape Z, using the time-stepping method to smooth/relax (43).
2. (Optional) Compute the AGC coefficients of the extended radiance model. Otherwise the AGC coefficients are the default ones (no AGC compensation).
3. Perform NV_f V-cycles of multigrid to solve for the optimal surface radiance f given a fixed shape Z, using the Jacobi or Gauss-Seidel methods as smoothing procedures on (39).

Figure 9: Multigrid method to solve the necessary optimality PDEs.

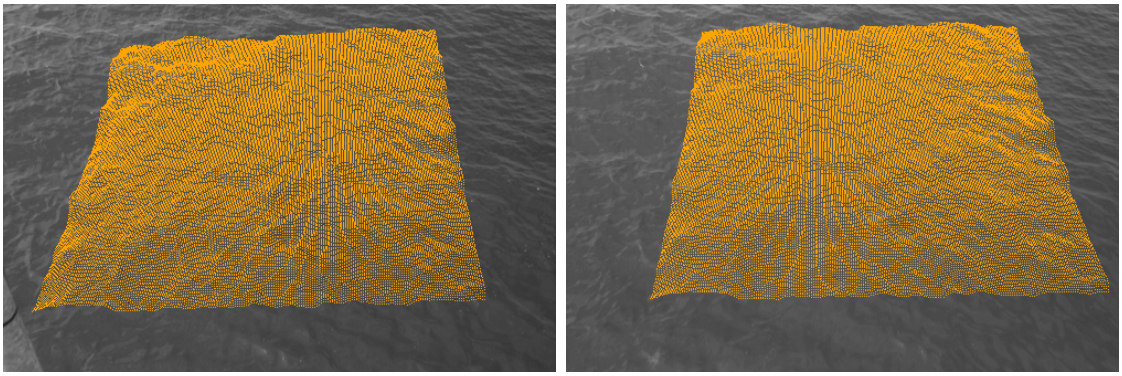


Figure 10: Projection of the estimated water surface on a stereo image. The graph or elevation map has been obtained on a grid of 210×210 points.

3.5 Experiments

3.5.1 Example 1. San Diego pier

The first reconstruction example to illustrate the elements involved in the proposed method is discussed next. Experiments with real data have been carried out with input stereo images of size 504×336 (Figure 8). The gradient descent PDEs are discretized on a 2-D grid with 210×210 points. The distance between grid points is 4 cm. Therefore, the grid covers an area of approximately $8.4 \times 8.4 m^2$. The deforming surface is initialized by the plane $Z = 0$, as shown in Figure 8. The time steps used in the experiments are a conservative measure (80%) of the approximate ℓ^2 stability time steps previously discussed.

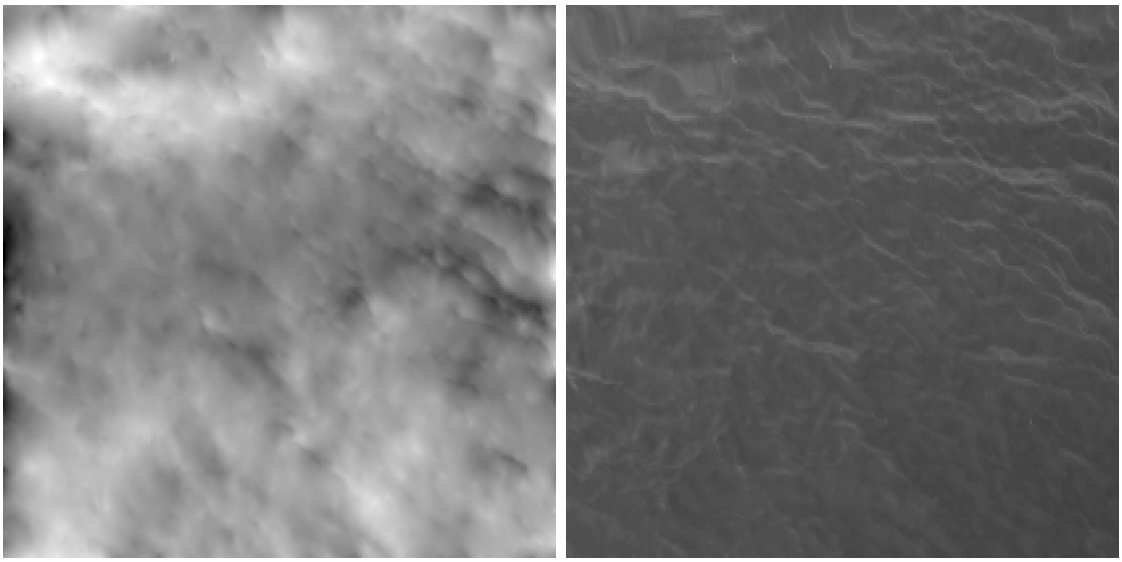


Figure 11: Left: estimated height function $Z(u, v)$ (shape of the water surface). Height is represented by greyscale intensities, from dark (low) to white (high). Right: estimated radiance function $f(u, v)$ (texture on the surface).

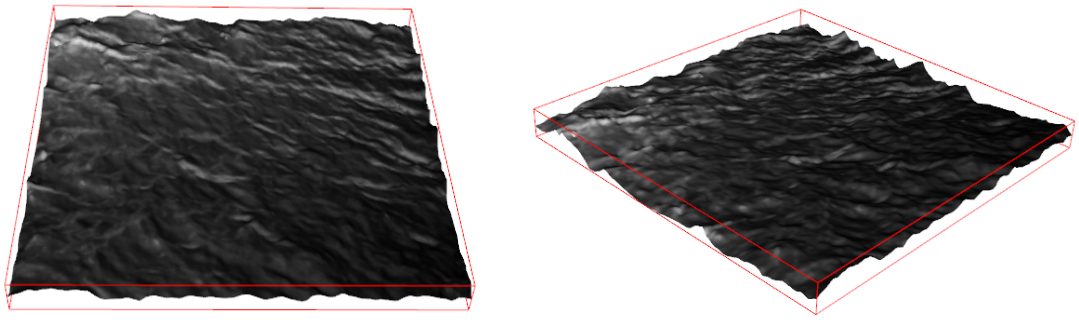


Figure 12: Perspective, three-dimensional representation of the estimated model of the water surface consisting of the shape (height) and radiance of the surface.

We first implemented our solution by means of a single-grid solver. Before using multi-grid methods with coarse-to-fine initialization (nested iteration [7]), the following updating scheme for the weights α and β of the regularizers was implemented to provide similar results but with slower convergence. The weights are initialized to high values so that the updates due to the diffusive gradient descent equations are first driven by the high smoothing terms and the first iterations capture the low frequency components of the solution. When the flows have converged, the weights are decreased (e.g., by a factor of 2) and the iteration proceeds. This scheme was controlled by an automatic criterion that used either

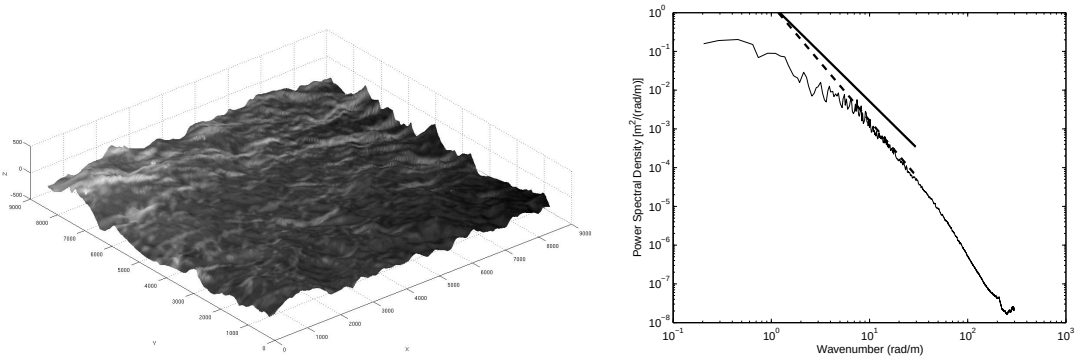


Figure 13: Left: orthographic three-dimensional representation of the estimated model of the water surface in MATLAB (cf. the right image in Figure 12). The vertical axis has been scaled to show details. Marks on axes are measured in millimeters. Right: Omni-directional wavenumber spectrum. Straight lines: $k^{-2.5}$ (solid) and k^{-3} (dashed).

the residual of the PDEs or the value of the energy to detect stagnation and overcome it. After approximately 10^5 iterations of gradient descent, the algorithm converges to the surface presented in Figure 10. With multigrid methods this updating scheme of the weights is no longer crucial due to the multiresolution nature of multigrid, and convergence is reduced to the order of hundreds of iterations (for this example). The final height and radiance functions are displayed on Figure 11. Both functions can be combined to generate a textured 3-D plot of the surface, as shown in Figure 12.

The reconstruction is validated beyond photo-metric criteria by known physical models that ocean waves must satisfy. In Figure 13, we report the omni-directional spectrum $S(k)$, where k is the wavenumber $k = \frac{2\pi}{\lambda}$ (rad/m), computed by integrating the directional spectrum $\Psi_Z(k, \theta)$ of the height function over all directions θ:

$$S(k) = \int_0^{2\pi} \Psi_Z(k, \theta) k \, \mathrm{d}\theta.$$

In agreement with turbulence theory [112, 94], the spectrum tail decays close to $k^{-2.5}$. Further, the empirical probability density functions derived from the reconstructed surface data compare very well with theoretical models [98, 22].

3.5.2 Example 2. Venice Canal

In another example, Figures 14 and 15 show a reconstructed water surface from images of the Venice Canal. Cropped images in Figure 14 are of size 600×450 pixels and show the

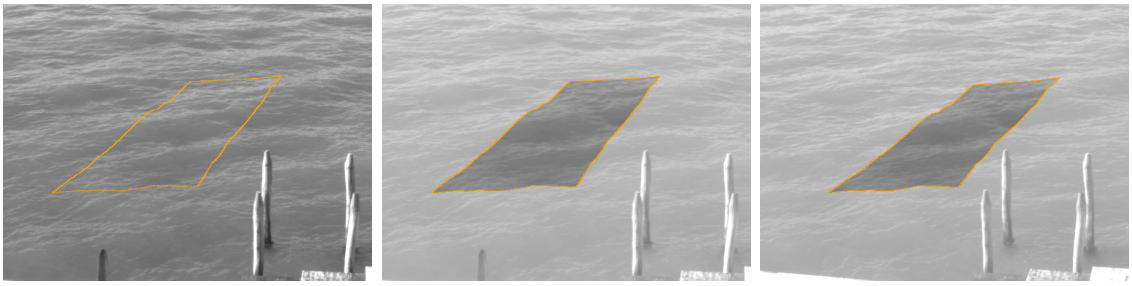


Figure 14: Left: projection on image 1 of the boundary of the estimated graph, which has been discretized by a grid of 129×513 points. Center: modeled image (computed form surface height and radiance) superimposed on original image 1. Right: modeled image 2 superimposed on original image 2.

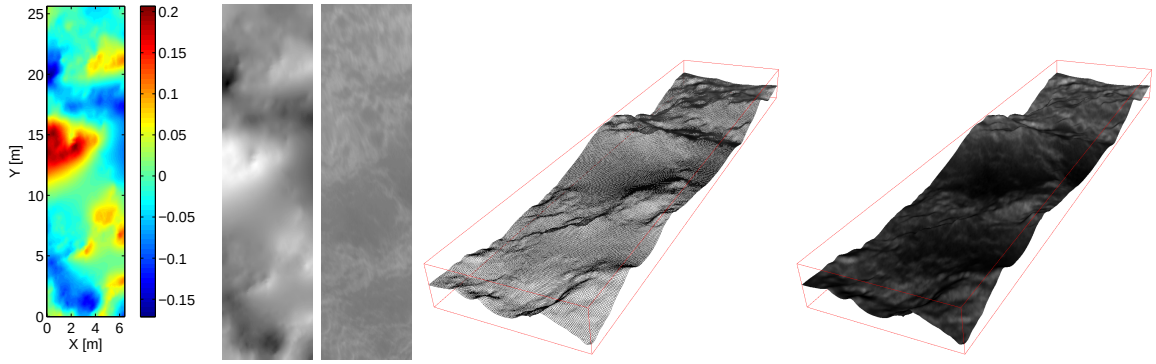


Figure 15: Form left to right: (1) estimated height function $Z(u, v)$ (shape of the water surface) in pseudo-color; (2) height function represented by grayscale intensities, from dark (low) to bright (high); (3) estimated radiance function $f(u, v)$ (texture on the surface); (4) perspective, three-dimensional wire-frame representation of the estimated surface shape (height) according to grid points; (5) texture-mapped surface obtained by incorporating the radiance function in the wire-frame model. In (4) and (5) the vertical axis has been magnified by a factor of 5 with respect to the horizontal axes for visualization purpose.

region of interest to be reconstructed. Figure 10 also displays the modeled images created by the generative model within our variational method. The data fidelity term compares the intensities of the original and modeled images in the highlighted region. As observed, the modeled images are a good match of the original images. Figure 15 shows the converged values of the unknowns of the problem: the height and the radiance of the surface, as well as the 3-D representation of the reconstructed surface obtained by combining both 2-D functions. In this experiment, the values of the weights of the regularizers were empirically determined: $\alpha = 0.035$ and $\beta = 0.01$. At the finest of the 5-level multigrid algorithm [7], the gradient descent PDEs are discretized on a 2-D grid with 129×513 points. The distance

between grid points is $h = 5$ cm. Therefore, the grid covers an area of $6.45 \times 25.65\, m^2$. An example of a surface discretized at the finest grid level is shown in Figure 15. Observe the high density of the surface representation, typical of variational methods. The step size h (distance between adjacent grid points) must be chosen so that it approximately matches the resolution in the images: a displacement of 1 pixel is observable at the finest grid level in the multigrid framework and it corresponds to a physical displacement of at least h. Due to perspective projection, the maximum value of h is determined by the grid points closest to the cameras.

3.6 Extended radiance model. Automatic Gain Control compensation

A common problem that appears when processing stereo images is that, even if the cameras that acquired the images of the scene are of the same model, the intensity values corresponding to the same scene point may systematically differ, that is, the Lambertian hypothesis on which the data fidelity term is based is not satisfied and (random) noise is not the cause. This may be due to different values of the internal, automatic gain control (AGC) of the cameras. Consequently, this systematic mismatch between the intensities in the images propagates within the algorithm that is designed to solve for the surface shape and radiance (by minimization of the energy) and degrades its performance. If the mismatch is not severe, the output of the algorithm may be close to the true solution, but as soon as the mismatch increases the output of the algorithm will be significantly different from the physical solution.

To overcome this systematic intensity error, we need to include additional parameters in the radiance model. The simple model that we use to correct the radiance of the surface states that the intensities at corresponding points in two distinct images differ according to a linear rule, which is completely described by two parameters: a scaling a and an offset b.

In a setup with N_c images, which are generated from a unique radiance function f, it might seem that we need to include the $2N_c$ parameters that relate the images to the generating radiance function, however this is not true. Given f and such a set of

$2N_c$ parameters $\{a_i, b_i\}_{i=1}^{N_c}$, the same generated images can be obtained by a linear transformation of f ($f' = Af + B$) and the inverse transformation of the $2N_c$ parameters: $\{a_i' = a_i/A,\, b_i' = b_i - a_i B/A\}_{i=1}^{N_c}$. Therefore, there are an infinite number of ways in which the parameters can be chosen to describe the same images. To have a unique set of parameters we would have to specify two additional constraints on f, such as that f has zero mean and unit energy. Thus, there are only $2N_c - 2$ degrees of freedom in the compensation model. To avoid enforcing constraints on f at each stage of an iterative solver, it is better to choose an image as the reference for f and relate the rest of the images to this one. This is equivalent to using $2N_c$ values $\{a_i, b_i\}_{i=1}^{N_c}$ in the AGC compensation model and setting $a_r = 1$ and $b_r = 0$, where r is the index of the reference image.

The latter model is adopted, where the photo-consistency criterion (15) in the data fidelity term, $\phi_i = \frac{1}{2}(I_i - f)^2$, is replaced by

$$\phi_i = \tfrac{1}{2}\big(I_i - (a_i f + b_i)\big)^2. \tag{55}$$

Now, the composite energy depends on two infinite-dimensional unknowns (Z and f) and a finite set of parameters $\{a_i, b_i\}_{i=2}^{N_c}$, which will be called the AGC parameters. Without loss of generality, assume that the first image is used as the reference. The necessary optimality condition for the new energy functional consists of the same EL equations as before (slightly modified) plus $N_c - 1$ equations, for $2 \le i \le N_c$, of the form

$$\frac{\partial J}{\partial a_i} = 0 = \frac{\partial J}{\partial b_i}.$$

There are several ways to compute the optimal AGC parameters. As with the case of the PDEs, one may opt for setting up gradient descent equations from some reasonable initialization values. Due to the finite-dimensional nature of the parameters, these equations will be ordinary differential equations (ODEs) and they will describe the evolution of the parameters as the height and radiance functions (Z and f, respectively) evolve in an alternating minimization approach. This is the general method to proceed for a generic

function $\phi_i(r)$ not necessarily quadratic in $r = I_i - (a_i f + b_i)$:

$$\begin{aligned}\frac{\partial a_i}{\partial t} &= -\frac{\partial J}{\partial a_i} \\ \frac{\partial b_i}{\partial t} &= -\frac{\partial J}{\partial b_i},\end{aligned}$$

where, since only the data-fidelity term depends on the AGC coefficients,

$$\begin{aligned}\frac{\partial J}{\partial a_i} &= \int_{\Omega_i} \dot{\phi}_i\big(I_i - (a_i f + b_i)\big)(-f)\,\mathrm{d}\hat{\mathbf{x}}_i, \\ \frac{\partial J}{\partial b_i} &= \int_{\Omega_i} \dot{\phi}_i\big(I_i - (a_i f + b_i)\big)(-1)\,\mathrm{d}\hat{\mathbf{x}}_i.\end{aligned}$$

Another option that is available in this case because ϕ_i is quadratic ($\dot{\phi}_i$ is linear) is to solve explicitly for the optimal AGC parameters at each stage of the alternating minimization approach, assuming Z and f are fixed. That is, for each image but the reference one,

$$\begin{aligned}\frac{\partial J}{\partial a_i} &= \int_{\Omega_i} \big(I_i - (a_i f + b_i)\big)(-f)\,\mathrm{d}\hat{\mathbf{x}}_i = 0, \\ \frac{\partial J}{\partial b_i} &= \int_{\Omega_i} \big(I_i - (a_i f + b_i)\big)(-1)\,\mathrm{d}\hat{\mathbf{x}}_i = 0.\end{aligned}$$

These form a linear system of two equations in two unknowns

$$\begin{pmatrix} \int_{\Omega_i} f^2\,\mathrm{d}\hat{\mathbf{x}}_i & \int_{\Omega_i} f\,\mathrm{d}\hat{\mathbf{x}}_i \\ \int_{\Omega_i} f\,\mathrm{d}\hat{\mathbf{x}}_i & \int_{\Omega_i} \mathrm{d}\hat{\mathbf{x}}_i \end{pmatrix} \begin{pmatrix} a_i \\ b_i \end{pmatrix} = \begin{pmatrix} \int_{\Omega_i} I_i f\,\mathrm{d}\hat{\mathbf{x}}_i \\ \int_{\Omega_i} I_i\,\mathrm{d}\hat{\mathbf{x}}_i \end{pmatrix}, \tag{56}$$

whose solution is straightforward to compute. Note, however, that the integrals are evaluated in the parameter space. Therefore, the actual system implemented is

$$\begin{pmatrix} \int_{\mathcal{U}} f^2 \mathtt{J}_i\,\mathrm{d}\mathbf{u} & \int_{\mathcal{U}} f \mathtt{J}_i\,\mathrm{d}\mathbf{u} \\ \int_{\mathcal{U}} f \mathtt{J}_i\,\mathrm{d}\mathbf{u} & \int_{\mathcal{U}} \mathtt{J}_i\,\mathrm{d}\mathbf{u} \end{pmatrix} \begin{pmatrix} a_i \\ b_i \end{pmatrix} = \begin{pmatrix} \int_{\mathcal{U}} I_i f \mathtt{J}_i\,\mathrm{d}\mathbf{u} \\ \int_{\mathcal{U}} I_i \mathtt{J}_i\,\mathrm{d}\mathbf{u} \end{pmatrix}.$$

How does the radiance model with AGC compensation affect the EL equations? In the case of the standard L^2 gradients in (37) and (39), only the terms derived from the data fidelity energy need to be modified:

$$\begin{aligned}\beta\Delta f + \sum_{k=1}^{N_c} a_k\big(I_k - (a_k f + b_k)\big)\mathtt{J}_k &= 0, \\ \alpha\Delta Z - \sum_{k=1}^{N_c} |\mathtt{M}^k| \tilde{Z}_k^{-3}\big(\hat{I}_k - (a_k \hat{f} + b_k)\big)(\mathbf{X} - \mathbf{C}_k)\cdot a_k \nabla \hat{f} &= 0.\end{aligned}$$

The incorporation of the AGC compensation in the height PDE is straightforward by replacing f with the new, extended radiance model $f \leftarrow a_i f + b_i$, and similarly with the derivative.

In the multigrid framework implemented for the numerical solution of the PDEs (Figure 9), the AGC coefficients are computed at the finest level available and are used in all coarser levels to maintain consistency of the model in the multiresolution setup.

3.7 Conclusion

In this chapter, a variational method for the three-dimensional reconstruction of surfaces defined in the form of a graph has been presented and discussed. For the aforementioned type of surfaces, this method has several advantages over methods designed for general surfaces, as analyzed in section 3.1. The method aims at the minimization of an energy functional that depends on the image data as well as some regularizing constraints on the unknowns to have a well-posed estimation problem. Calculus of Variations provides the necessary optimality conditions to search for the minimizer. These comprise a system of coupled PDEs, whose mathematical rigorous derivation has been moved to the appendices to maintain focus of the thesis. A numerical solution of the PDE system is sought using finite differences and an iterative alternating minimization approach. Based on the complexity of several numerical methods discussed to solve the easiest of the "decoupled" PDEs, the election of multigrid methods has been justified and has likewise been adopted for the nonlinear PDE. Examples of the proposed method applied to the reconstruction of oceanic sea states have also been given to guide the reader in the understanding of the different elements in the method, e.g., images, height function, radiance function, reconstructed surface, etc. Finally, a minor improvement of the method has been proposed and discussed via an extended radiance model to address the practical problem of cameras with different gains, the so-called AGC correction or compensation.

CHAPTER IV

SCENE PLANE ESTIMATION VIA INDUCED 2-D HOMOGRAPHY

For the variational graph algorithm introduced in chapter 3 to work, it is necessary to specify a reference plane in Euclidean space with respect to which the graph is defined. In an ideal case, this plane is the one aligned in the direction of gravity acceleration (an horizontal plane), thus conferring the meaning of *height* or *elevation map* to the graph. In the images of the ocean waves under consideration, there is no known reference that allow a reliable estimation of the vertical direction. Thus, it is a challenging problem to be able to estimate this vertical direction solely from the images of the ocean and the frame in which the cameras are given (usually, one of the cameras is chosen to be the reference for the world coordinates). Equivalently stated, it is challenging to be able to transform coordinates from the frame in which the cameras are given into a frame adapted to the waves: horizontal plane and vertical direction. In a scenario with measurements (images) over a long period of time, the gravity direction can be found by exploiting the hypothesis that the mean time water elevation map is a plane [4] (representing the water surface in a calmed state) orthogonal to gravity direction.

Consider the scenario of a calmed ocean surface with almost no waves observed at a particular time. One approach to estimate the gravity plane is to compute the best plane that fits the reconstructed ocean waves when the surface is represented as a collection of scattered 3-D points. Several methods falling in this category are discussed in appendix G; they only use geometric information of the surface. However, this approach might fail if there are no strong features on the "texture" representing the surface of the ocean and thus the surface cannot be accurately reconstructed. Another possible approach is to exploit the flatness assumption and obtain the plane by optimization of an aggregate photometric cost function that measures how well a candidate plane through the scene explains the observed images. This is the method developed in this chapter, in a setup with two cameras.

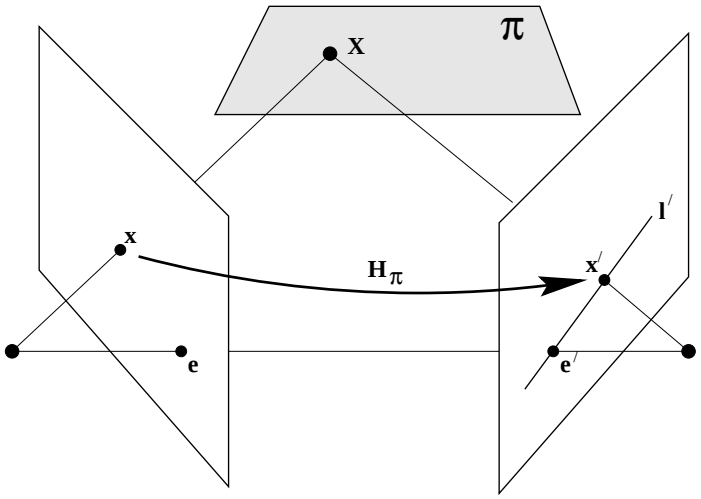


Figure 16: A plane $\boldsymbol{\pi}$ in space induces a 2-D homography between two image planes.

4.1 Minimization of symmetrically transfered photometric error

We consider the problem of estimating the plane in the scene that minimizes an aggregate photometric error. Let the plane in the scene have homogeneous coordinates $\boldsymbol{\pi} = (a, b, c, d)^\top$. This plane[1] induces a correspondence between the two images viewing the plane [18, 35, 59], as shown in Figure 16. This mapping is a homography, i.e., a linear projective transformation, more precisely, a *planar homography* [18, p.270], and it is denoted by $\mathtt{H}(\boldsymbol{\pi})$. In particular, let $\mathtt{H}$ represent the homography from image 1 to image 2. We propose to find the plane $\boldsymbol{\pi}$ that minimizes

$$J(\boldsymbol{\pi}) = J_1(\boldsymbol{\pi}) + J_2(\boldsymbol{\pi}), \tag{57}$$

where the photometric discrepancy between an image and its transferred version through the homography $\mathtt{H}$ is measured in the L^2 sense, in both images, by:

$$J_1(\boldsymbol{\pi}) = \int_\Omega \tfrac{1}{2}\Big(I(\mathbf{x}_e) - I'\big((\mathtt{H}\mathbf{x})_e\big)\Big)^2 \,\mathrm{d}\mathbf{x}_e, \tag{58}$$

$$J_2(\boldsymbol{\pi}) = \int_{\Omega'} \tfrac{1}{2}\Big(I\big((\mathtt{H}^{-1}\mathbf{x}')_e\big) - I'\big(\mathbf{x}'_e\big)\Big)^2 \,\mathrm{d}\mathbf{x}'_e. \tag{59}$$

Quantities in image 2 are represented with a prime and quantities in image 1, without a prime. The subscript $\cdot_e$ emphasizes the fact that Euclidean coordinates are used to index

[1] The plane must not go through either of the two optical centers of the cameras.

the images, even though homogeneous coordinates may be used to denote the action of the homography on image points. The simplest integration regions Ω and Ω' are interior rectangles in the images (located away from the edges of the images by some margin). It is clear that (57) is symmetric with respect to $\mathtt{H}$: it considers the error measured in both images by transferring image intensities through $\mathtt{H}$ (from image 1 to image 2) and $\mathtt{H}^{-1}$ (from image 2 to image 1) .

The 3×3 matrix of the homography $\mathtt{H}$ is determined by the projection matrices of the cameras associated to both images and by the coordinates of the plane $\boldsymbol{\pi}$ (in the same world frame as the projection matrices). Let the projection matrices be $\mathtt{P} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3)^\top$ and $\mathtt{P}' = (\boldsymbol{\pi}'_1, \boldsymbol{\pi}'_2, \boldsymbol{\pi}'_3)^\top$. The homogeneous matrix of the homography is, following the formula in [18, p.253], up to a non-zero scale factor,

$$\mathtt{H}(\boldsymbol{\pi}, \mathtt{P}, \mathtt{P}') \sim \begin{pmatrix} \det(\boldsymbol{\pi}'_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3, \boldsymbol{\pi}) & \det(\boldsymbol{\pi}'_1, \boldsymbol{\pi}_3, \boldsymbol{\pi}_1, \boldsymbol{\pi}) & \det(\boldsymbol{\pi}'_1, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}) \\ \det(\boldsymbol{\pi}'_2, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3, \boldsymbol{\pi}) & \det(\boldsymbol{\pi}'_2, \boldsymbol{\pi}_3, \boldsymbol{\pi}_1, \boldsymbol{\pi}) & \det(\boldsymbol{\pi}'_2, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}) \\ \det(\boldsymbol{\pi}'_3, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3, \boldsymbol{\pi}) & \det(\boldsymbol{\pi}'_3, \boldsymbol{\pi}_3, \boldsymbol{\pi}_1, \boldsymbol{\pi}) & \det(\boldsymbol{\pi}'_3, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}) \end{pmatrix}, \tag{60}$$

that is, $\mathtt{H}(\boldsymbol{\pi}, \mathtt{P}, \mathtt{P}') \sim (h_{ij})$, where

$$h_{ij} = \det(\boldsymbol{\pi}'_i, \boldsymbol{\pi}_{j_1}, \boldsymbol{\pi}_{j_2}, \boldsymbol{\pi}) \tag{61}$$

and $j \mapsto (j_1, j_2)$ is specified by $1 \mapsto (2,3)$, $2 \mapsto (3,1)$, $3 \mapsto (1,2)$. Moreover, swapping the roles of the projection matrices in the formula for $\mathtt{H}$ gives the formula for its inverse: $\mathtt{H}^{-1}(\boldsymbol{\pi}, \mathtt{P}, \mathtt{P}') \sim \mathtt{H}(\boldsymbol{\pi}, \mathtt{P}', \mathtt{P})$.

The unknown in the cost (57) is finite-dimensional (a 4-vector) and the cost does not depend on the derivatives of the plane $\boldsymbol{\pi}$, therefore the strategy followed to optimize (57) consists of applying a standard minimization algorithm that uses information of the first derivative, e.g., steepest descent or conjugate gradient (Fletcher-Reeves and Polak-Riviere variations) [76]. To apply Newton's method one would need to provide the second order derivatives (Hessian) of (57), but it is analytically difficult to derive those derivatives and computationally expensive. First order derivatives are enough for this problem.

4.1.1 Analytical expression for the gradient of the cost function

Let us find an analytical expression for the first derivative (i.e., gradient) of $J_1(\boldsymbol{\pi})$:

$$(\nabla J_1)^\top \equiv \frac{\partial J_1}{\partial \boldsymbol{\pi}} = \int_\Omega \Big(I'\big((\mathtt{H}\mathbf{x})_e - I(\mathbf{x}_e)\big)\Big) \frac{\partial I'\big((\mathtt{H}\mathbf{x})_e\big)}{\partial \boldsymbol{\pi}} \, \mathrm{d}\mathbf{x}_e. \tag{62}$$

Applying the chain rule, letting $\nabla I' = (I'_x, I'_y)^\top$ be the standard image gradient and letting $\mathbf{x}' = \mathtt{H}\mathbf{x}$, with $\mathbf{x} = (x, y, z)^\top$, leads to

$$\frac{\partial I'\big(\mathbf{x}'_e\big)}{\partial \boldsymbol{\pi}} = \nabla I'^\top \frac{\partial \mathbf{x}'_e}{\partial \mathbf{x}'} \frac{\partial \mathbf{x}'}{\partial \boldsymbol{\pi}}. \tag{63}$$

Let $\mathbf{x}' = (x', y', z')^\top$, then $\mathbf{x}'_e = (x'_e, y'_e)^\top$, with $x'_e = x'/z'$, $y'_e = y'/z'$, and

$$\frac{\partial \mathbf{x}'_e}{\partial \mathbf{x}'} = \frac{1}{z'} \begin{pmatrix} 1 & 0 & -x'_e \\ 0 & 1 & -y'_e \end{pmatrix}.$$

The last derivative in the chain rule (63) is the 3×4 matrix

$$\frac{\partial \mathbf{x}'}{\partial \boldsymbol{\pi}} = \frac{\partial (\mathtt{H}\mathbf{x})}{\partial \boldsymbol{\pi}} = \begin{pmatrix} x\partial_{\boldsymbol{\pi}} h_{11} + y\partial_{\boldsymbol{\pi}} h_{12} + z\partial_{\boldsymbol{\pi}} h_{13} \\ x\partial_{\boldsymbol{\pi}} h_{21} + y\partial_{\boldsymbol{\pi}} h_{22} + z\partial_{\boldsymbol{\pi}} h_{23} \\ x\partial_{\boldsymbol{\pi}} h_{31} + y\partial_{\boldsymbol{\pi}} h_{32} + z\partial_{\boldsymbol{\pi}} h_{33} \end{pmatrix}. \tag{64}$$

Each entry in (64) is the sum of three terms of the form $w\partial_{\boldsymbol{\pi}} h_{ij}$, with h_{ij} as in (61). Let $\mathtt{B} = (\mathbf{u}, \mathbf{v}, \mathbf{w}, \boldsymbol{\pi})$ be the 4×4 matrix whose determinant is an arbitrary entry $h_{ij} = \det(\mathtt{B})$. If $\mathbf{e}_k$ is the k-th canonical basis vector in $\mathbb{R}^4$, one can show that

$$\frac{\partial h_{ij}}{\partial \boldsymbol{\pi}} = \det(\mathtt{B})\, \mathbf{e}_4^\top \mathtt{B}^{-1} = \big(\mathrm{cof}(\mathtt{B})\mathbf{e}_4\big)^\top,$$

which is the 4-th column (in row form) of the cofactor matrix in $\mathtt{B}^{-1} = \det^{-1}(\mathtt{B})(\mathrm{cof}(\mathtt{B}))^\top$. It turns out that such column solely depends on the first three columns of $\mathtt{B}$, i.e., it does not depend on $\boldsymbol{\pi}$. To show this, given a 3×4 matrix $\mathtt{G} = (\mathbf{u}, \mathbf{v}, \mathbf{w})^\top$ define the following vector in its null-space:

$$\ker(\mathtt{G}) = (D_{234,} - D_{134}, D_{124}, -D_{123})^\top, \tag{65}$$

where D_{ijk} is the minor of $\mathtt{G}$ obtained by choosing columns i, j, k. Then $\operatorname{cof}(\mathtt{B})\mathbf{e}_4 = -\ker\big((\mathbf{u},\mathbf{v},\mathbf{w})^\top\big)$ and (64) becomes (clearly independent of $\boldsymbol{\pi}$)

$$\frac{\partial \mathbf{x}'}{\partial \boldsymbol{\pi}} = -\begin{pmatrix} x\ker^\top\big((\boldsymbol{\pi}'_1,\boldsymbol{\pi}_2,\boldsymbol{\pi}_3)^\top\big) + y\ker^\top\big((\boldsymbol{\pi}'_1,\boldsymbol{\pi}_3,\boldsymbol{\pi}_1)^\top\big) + z\ker^\top\big((\boldsymbol{\pi}'_1,\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)^\top\big) \\ x\ker^\top\big((\boldsymbol{\pi}'_2,\boldsymbol{\pi}_2,\boldsymbol{\pi}_3)^\top\big) + y\ker^\top\big((\boldsymbol{\pi}'_2,\boldsymbol{\pi}_3,\boldsymbol{\pi}_1)^\top\big) + z\ker^\top\big((\boldsymbol{\pi}'_2,\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)^\top\big) \\ x\ker^\top\big((\boldsymbol{\pi}'_3,\boldsymbol{\pi}_2,\boldsymbol{\pi}_3)^\top\big) + y\ker^\top\big((\boldsymbol{\pi}'_3,\boldsymbol{\pi}_3,\boldsymbol{\pi}_1)^\top\big) + z\ker^\top\big((\boldsymbol{\pi}'_3,\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)^\top\big) \end{pmatrix}. \quad (66)$$

Moreover, since $\det(\mathbf{u},\mathbf{v},\mathbf{w},\boldsymbol{\pi}) = \boldsymbol{\pi}^\top \operatorname{cof}_4(\mathbf{u},\mathbf{v},\mathbf{w},\boldsymbol{\pi})$, it follows that (61) is also

$$h_{ij} = -\boldsymbol{\pi}^\top \ker\big((\boldsymbol{\pi}'_i,\boldsymbol{\pi}_{j_1},\boldsymbol{\pi}_{j_2})^\top\big). \quad (67)$$

An alternative formula for (60) is, according to [18],

$$\mathtt{H} \sim \mathtt{P}'\mathtt{P}^+_{\boldsymbol{\pi}}, \quad (68)$$

where $\mathtt{P}^+_{\boldsymbol{\pi}} = (\boldsymbol{\pi}_2 \triangle \boldsymbol{\pi}_3 \triangle \boldsymbol{\pi}, \boldsymbol{\pi}_3 \triangle \boldsymbol{\pi}_1 \triangle \boldsymbol{\pi}, \boldsymbol{\pi}_1 \triangle \boldsymbol{\pi}_2 \triangle \boldsymbol{\pi})$, and $\mathbf{u} \triangle \mathbf{v} \triangle \mathbf{w} = \ker\big((\mathbf{u},\mathbf{v},\mathbf{w})^\top\big) = \mathtt{M}^*(\mathbf{u},\mathbf{v})\mathbf{w}$, with $\mathtt{M}^*(\mathbf{u},\mathbf{v})$ being a 4×4 skew-symmetric matrix [78]. Since $\frac{\partial}{\partial \mathbf{w}}(\mathbf{u}\triangle\mathbf{v}\triangle\mathbf{w}) = \mathtt{M}^*(\mathbf{u},\mathbf{v})$, the last derivative in (63) is the 3×4 matrix

$$\frac{\partial \mathbf{x}'}{\partial \boldsymbol{\pi}} = \frac{\partial}{\partial \boldsymbol{\pi}}(\mathtt{P}'\mathtt{P}^+_{\boldsymbol{\pi}}\mathbf{x}) = \mathtt{P}'\left(x\mathtt{M}^*(\boldsymbol{\pi}_2,\boldsymbol{\pi}_3) + y\mathtt{M}^*(\boldsymbol{\pi}_3,\boldsymbol{\pi}_1) + z\mathtt{M}^*(\boldsymbol{\pi}_1,\boldsymbol{\pi}_2)\right). \quad (69)$$

Because (63) has to be computed at every point of the region of interest Ω and for many iterations of the optimization method, it is worth counting the operations involved to choose the best expression for (64). The implementation of (61) and (66) requires 9 evaluations of (65), while the implementation of (68) and (69) requires only 3 evaluations, 3 skew-symmetric matrices and a matrix-matrix product. Therefore, the latter is faster.

Similar formulas apply to the derivative if the inverse is used, $\mathtt{H}^{-1} \sim \mathtt{P}\mathtt{P}'^+_{\boldsymbol{\pi}}$, as in J_2. Hence, the gradient of (57) is $\nabla J = \nabla J_1 + \nabla J_2$, with

$$(\nabla J_2)^\top = \int_{\Omega'} \Big(I\big((\mathtt{H}^{-1}\mathbf{x}')_e\big) - I'(\mathbf{x}'_e)\Big)\, \frac{\partial I\big((\mathtt{H}^{-1}\mathbf{x}')_e\big)}{\partial \boldsymbol{\pi}}\, \mathrm{d}\mathbf{x}'_e, \quad (70)$$

where $\mathbf{x} = \mathtt{H}^{-1}\mathbf{x}'$ and

$$\frac{\partial I\big(\mathbf{x}_e\big)}{\partial \boldsymbol{\pi}} = \nabla I^\top \frac{\partial \mathbf{x}_e}{\partial \mathbf{x}} \frac{\partial(\mathtt{H}^{-1}\mathbf{x}')}{\partial \boldsymbol{\pi}}. \quad (71)$$

Clearly, ∇J_1 and ∇J_2 have similar expressions due to the symmetry in the definition of J_1 and J_2: (70) can be ontained from (62) by swapping $I \leftrightarrow I'$, $\mathbf{x} \leftrightarrow \mathbf{x}'$ and $\mathtt{H} \leftrightarrow \mathtt{H}^{-1}$.

4.2 Numerical considerations

Observe that the computation of (61) involves sums of 4-th order products of coordinates. The same applies to the derivative term (66). If there is a significant difference in the order of magnitude of the entries of $\mathtt{P}$, $\mathtt{P}'$ and $\boldsymbol{\pi}$, data normalization [36] must be carried out prior to applying the formulas in the previous sections to avoid a loss in numerical precision. The formula that relates the induced homographies before and after normalization of the coordinates in the image $\mathtt{T}$ and in space $\mathtt{U}$ is

$$\mathtt{H}(\mathtt{U}^{\top}\boldsymbol{\pi}, \mathtt{TPU}, \mathtt{TP}'\mathtt{U}) = \det(\mathtt{U})\det(\mathtt{T})\,\mathtt{T}\,\mathtt{H}(\boldsymbol{\pi}, \mathtt{P}, \mathtt{P}')\,\mathtt{T}^{-1}. \tag{72}$$

To improve the numerical performance of the implementation in finite precision of the iterative minimization algorithm, it is necessary to find transformations $\mathtt{T}$ (3×3) and $\mathtt{U}$ (4×4) such that the left hand side of (72) is better numerically conditioned than $\mathtt{H}(\boldsymbol{\pi}, \mathtt{P}, \mathtt{P}')$, and such that the derivatives are also more stable numerically to avoid loss of precision. Assuming the projection matrices $\mathtt{P}$ and $\mathtt{P}'$ are given in a Euclidean frame (and so is the plane $\boldsymbol{\pi}$), the simplest matrices $\mathtt{T}$ and $\mathtt{U}$ that yield the desired effect are those of similarity transformations. Let $\mathtt{T}$ be, as is customary, a translation and uniform scaling of the image coordinates so that the image center becomes the origin and the transformed coordinates are approximately in the $[-1, 1]^2$ square (or same order of magnitude):

$$\mathtt{T} = \begin{pmatrix} \frac{2}{w+h} & 0 & -\frac{w}{w+h} \\ 0 & \frac{2}{w+h} & -\frac{h}{w+h} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} s_{\mathtt{T}}\mathtt{I} & \mathbf{t}_{\mathtt{T}} \\ \mathbf{0}^{\top} & 1 \end{pmatrix}, \tag{73}$$

where w and h are the width and height of the image(s) I (and I'), respectively. This transfomation allows to compensate for large differences in the orders of magnitude of the rows of $\mathtt{P} = \mathtt{K}[\mathtt{R} \mid \mathbf{t}]$ and $\mathtt{P}' = \mathtt{K}'[\mathtt{R}' \mid \mathbf{t}']$. A uniform scaling of the world coordinates (without rotation or translation) allows $\mathtt{U}$ to overcome large differences in the orders of magnitude of the columns of the projection matrices: $\mathtt{U} = \mathrm{diag}(s_u, s_u, s_u, 1)$, where $s_u = (||\mathbf{t}|| + ||\mathbf{t}'||)/2$ is a reasonable value for the scaling factor.

Notice that, in (57), the parameter vector $\boldsymbol{\pi}$ (a plane in space) is solely affected by the space similarity $\mathtt{U}$, not by the image similarity $\mathtt{T}$. The space normalization does not affect

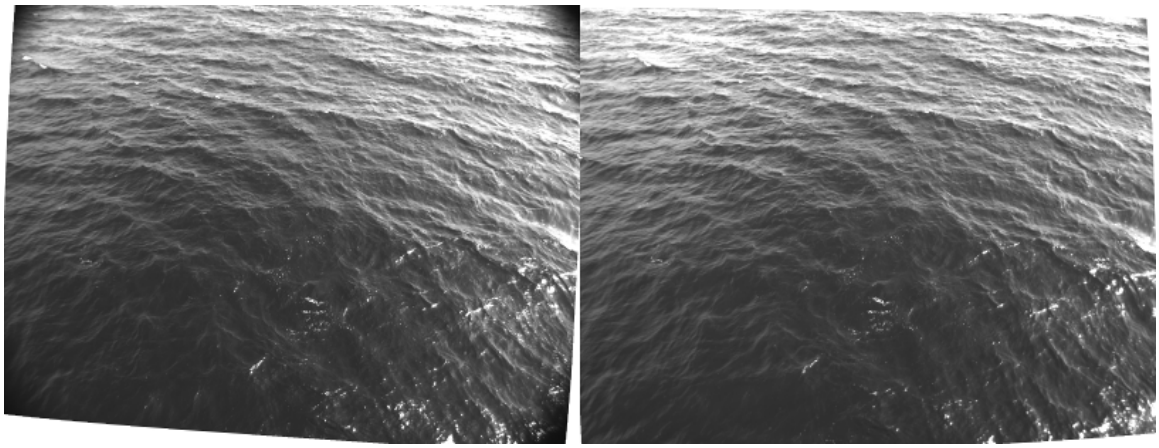


Figure 17: A (stereo) pair of images. Left: $I(\mathbf{x})$, $\mathbf{x} \in \Omega$. Right $I'(\mathbf{x}')$, $\mathbf{x}' \in \Omega'$.

the cost function J, however it does change the expression of ∇J. The image normalization improves the precision in the numerical evaluation of ∇J without changing its formula.

The previous approach has some limitations. The accuracy of the composite result $\frac{\partial \mathbf{x}'}{\partial \boldsymbol{\pi}}$ can be spoiled by the magnification factor within $\mathtt{T}^{-1}$ ($s_{\mathtt{T}}^{-1}$ in the order of thousands of pixels). The second factor in the chain rule (63) is also carried out in unnormalized coordinates and this is undesirable. To overcome these issues, the analytic expression for the gradient should be derived directly in normalized image/camera coordinates, as it has been implemented.

4.3 Discussion

There is no need to write special optimization software for this problem. General purpose implementations of the conjugate gradient method are publicly available in many C++ libraries, for example, the GNU-GSL v1.14 library. The cost and gradient functions passed to the optimization method evaluate an approximation of integrals (58), (59), (62) and (70), on discrete grids in the image domains Ω and Ω'. Image derivatives at a given location are computed using the analytical gradient of a bilinear function that fits the intensities of the four closest pixels to the given location.

Figures 17 through 21 show an example of the application of the proposed method to estimate the plane in the scene that best explains a stereo pair of images of ocean waves. The size of each image in the stereo pair is 1624×1236 pixels. Regions of interest (ROIs)

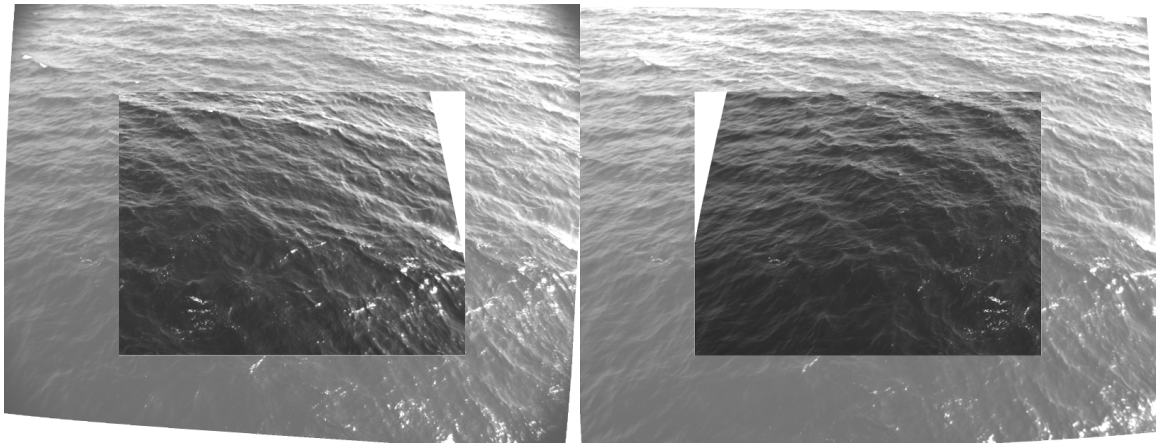


Figure 18: Before optimization. Modeled ROIs with intensities borrowed from the original images. Left: $\hat{I}(\mathbf{x}) = I'(\mathtt{H}\mathbf{x})$, $\mathbf{x} \in \Omega$. Right: $\hat{I}'(\mathbf{x}') = I(\mathtt{H}^{-1}\mathbf{x}')$, $\mathbf{x}' \in \Omega'$, with $\mathtt{H} = \mathtt{H}(\boldsymbol{\pi}_0)$.

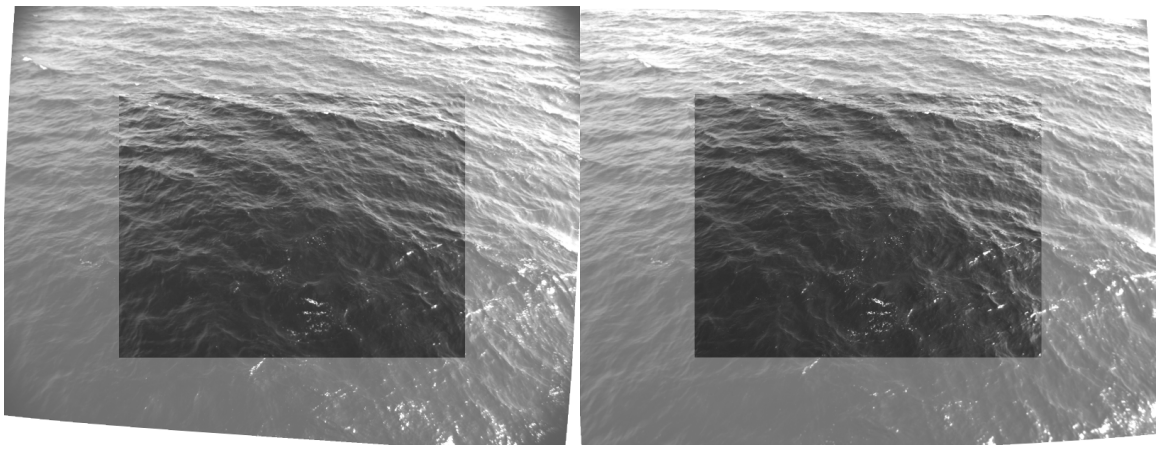


Figure 19: After optimization. Modeled ROIs with intensities borrowed from the original images. Left: $\hat{I}(\mathbf{x}) = I'(\mathtt{H}\mathbf{x})$, $\mathbf{x} \in \Omega$. Right: $\hat{I}'(\mathbf{x}') = I(\mathtt{H}^{-1}\mathbf{x}')$, $\mathbf{x}' \in \Omega'$, with $\mathtt{H} = \mathtt{H}(\boldsymbol{\pi}_{\min})$, where $\boldsymbol{\pi}_{\min}$ is the minimizer of the cost function given by a line search optimization method.

centered at the images are chosen as integration domains Ω and Ω' (see Figure 19). For a given initial candidate plane $\boldsymbol{\pi}_0$, the corresponding modeled ROIs are shown in Figure 18. These modeled ROIs are clearly not a good match to the original images in Figure 17. Upon convergence of the method, the solution plane provides the much better fit displayed in Figure 19. The previous modeled images are the ones used in the discretization of the cost function. However, we can also show the transferred ROIs through the action of the homographies corresponding to the initial and final planes, as displayed in Figures 20 and 21, respectively. The transformation of the integration domains Ω and Ω' is evident. Again, the proposed method significantly improves the photometric match between the modeled

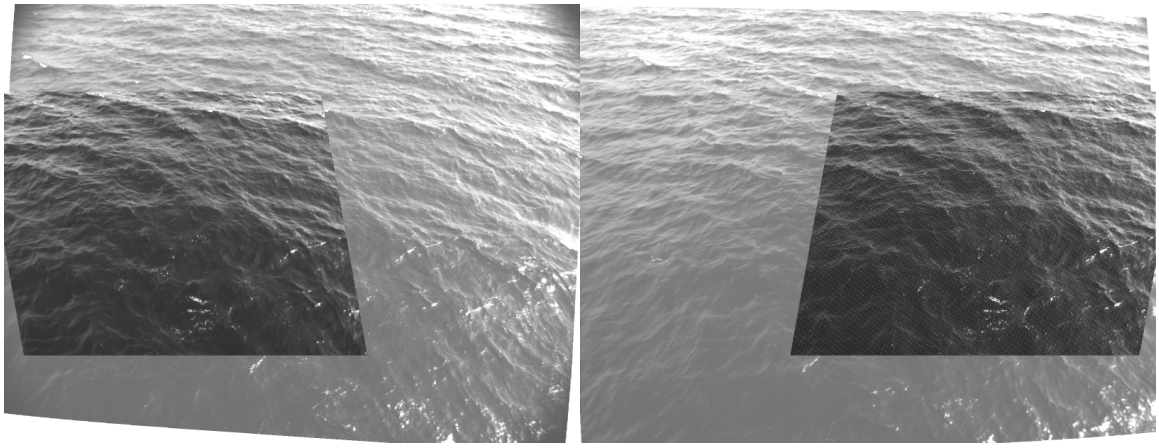


Figure 20: Before optimization. Transferred ROIs. Left: $\hat{I}(\mathtt{H}^{-1}\mathbf{x}') = I'(\mathbf{x}')$, $\mathbf{x}' \in \Omega'$, i.e., $\mathtt{H}^{-1}\Omega'$. Right: $\hat{I}'(\mathtt{H}\mathbf{x}) = I(\mathbf{x})$, $\mathbf{x} \in \Omega$, i.e., $\mathtt{H}\Omega$, with $\mathtt{H} = \mathtt{H}(\boldsymbol{\pi}_0)$.

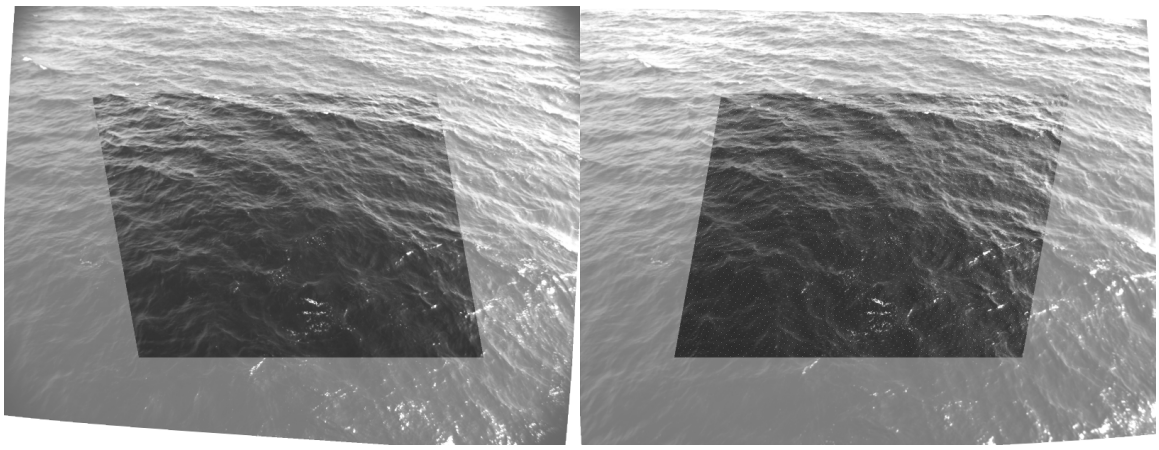


Figure 21: After optimization. Transferred ROIs. Left: $\hat{I}(\mathtt{H}^{-1}\mathbf{x}') = I'(\mathbf{x}')$, $\mathbf{x}' \in \Omega'$, i.e., $\mathtt{H}^{-1}\Omega'$. Right: $\hat{I}'(\mathtt{H}\mathbf{x}) = I(\mathbf{x})$, $\mathbf{x} \in \Omega$, i.e., $\mathtt{H}\Omega$, with $\mathtt{H} = \mathtt{H}(\boldsymbol{\pi}_{\min})$.

and the original images given by the initial plane $\boldsymbol{\pi}_0$

To achieve successful results, the non-linear optimization algorithm must have a good initialization. The non-photometric algorithms of appendix G can serve this purpose because they only require a few data points to provide a candidate plane $\boldsymbol{\pi}$. Another possible method to obtain an initial plane relies on exploiting the following bidirectional property: for planes in general position, the induced homography is determined uniquely by the plane, and *vice versa* [35]. Therefore, one may estimate a general 2-D homography [35] from feature points in the images (if there is enough texture), "project" this mapping onto the class of planar homographies, and retrieve the associated plane.

The proposed method is used for initialization of the variational graph method because the latter requires the determination of a plane (ideally, the gravity plane) with respect to which the surface is represented in the form of a graph. It has been noticed that the application of this algorithm to the initialization of the surface evolution has beneficial effects. For the same number of multigrid iterations, the composite energy of the variational graph method is smaller than that obtained if the surface is defined with respect to planes computed by non-photometric means (section G). This is both due to a better initial estimate of the surface (if initialized by the zero plane) and a better representation of the surface, both resulting in a smaller penalty due to the geometric regularizer. The wave height distribution presents less asymmetry (skewness) when using this plane and the maximum crest-to-trough height also decreases. In addition, the method has another application: it provides a (flat) surface that can serve as a reference for comparison of the reconstructed surface by means of the variational graph method. Since the method proposed in this chapter is mainly used for initialization of the variational graph method, i.e., it is not the core of the thesis, we do not pursue a further detailed analysis of the performance of the aforementioned method.

CHAPTER V

TOWARD INCORPORATING PHYSICS OF THE SURFACE IN THE VARIATIONAL RECONSTRUCTION METHOD

So far, the regularizing terms (19) and (20) have no physical meaning according to the dynamics of the water waves. They are the simplest smoothness penalties to support the conjecture that the problem is well posed and a solution exists, without providing a rigorous proof. In the next sections some discussion about other choices for these terms will be given.

5.1 Area-based regularizer

Since the regularizer on the shape of the surface (19) acts on a geometric object, a more sensible geometric choice that does not significantly complicate the model is to penalize the total area of the reconstructed surface:

$$E_{\text{geom}} = \int_S \mathrm{d}A = \int_U \sqrt{1 + Z_u^2 + Z_v^2}\,\mathrm{d}\mathbf{u}. \tag{74}$$

Surfaces that minimize the above energy are called minimal surfaces and have the property of zero mean curvature. As explained in appendix C.3, if (74) is used in (12), the diffusive term in the PDE (37), i.e., the Laplacian $\Delta Z = Z_{uu} + Z_{vv}$, is replaced by the mean curvature

$$2H = \frac{(1 + Z_v^2)Z_{uu} - 2Z_u Z_v Z_{uv} + (1 + Z_u^2)Z_{vv}}{(1 + Z_u^2 + Z_v^2)^{3/2}}. \tag{75}$$

Calculations show that the new regularizer does not alter the homogeneous Neumann boundary condition. Moreover, if the gradient of the functional is defined with respect to the standard *geometric* L^2 inner product (not the regular L^2), the gradient descent PDE in the height is replaced by (see appendix C.3) the geometric law:

$$Z_t = \frac{\alpha 2H - g(Z)}{\sqrt{1 + Z_u^2 + Z_v^2}}. \tag{76}$$

This model is slightly more involved than (43) due to the more complicated expression for the mean curvature diffusive term (compared to the Laplacian) and the square root in the denominator. However the square root only needs to be carried out for the non-linear term $g(Z)$ since no square root appears in the diffusive term due to cancellation with the denominator of (75). The reason why a more geometric regularizer for the radiance is not proposed is to avoid altering the radiance PDE and complicating the model since the radiance is just an auxiliary variable to better drive the evolution of the surface shape. Recall tha if $\beta = 0$ the optimal radiance is a weighted image average and no radiance regularizer is needed at all to estimate the surface shape (but in this case the solver is prone to get stuck more easily in local minima). Therefore, we still consider the simple and powerful PDEs (39)-(40) to describe the radiance problem.

Assuming the explicit FTCS updating scheme is used to relax the modified non-linear PDE in the height, an ℓ^2 stability condition for the time step can be derived using Fourier analysis under reasonable approximations, as shown in appendix F.4:

$$0 \leq \frac{\alpha}{h^2} \leq \min \dot{g}(\bar{Z}_{i,j}^n) \tag{77}$$

and

$$\Delta t \leq \big(\frac{1}{2} \max |\dot{g}(\bar{Z}_{i,j}^n)| + \frac{5\alpha}{h^2}\big)^{-1}. \tag{78}$$

The maximum time step Δt has the same form as (54), but with $4\alpha/h^2$ replaced by $5\alpha/h^2$. Conditions (77) and (78), derived from the approximate Fourier analysis in the worst case scenario, are sufficient to guarantee stability in ℓ^2 of the time-stepping, finite difference scheme. The scheme may be stable even if these conditions are not satisfied.

In practice, it is observed that the results obtained with the newly proposed regularizer in the surface shape are just a minor modification of the results obtained with the simple L^2 regularizer (19) (for the same value of the weight α in (12)). This will be measured in the upcoming chapter on experimental validation. The statistical regularizers described next have a larger impact on the surface shape than (74).

5.2 Regularizers on the statistics of the surface

The flexibility of the variational framework allows us to incorporate properties of the physics of the waves in the model that would otherwise be difficult to account for via other stereo methods. For example, we may include global statistical properties in the form of a weak constraint by considering an extra energy term that penalizes the deviation of the statistics of the reconstructed surface with respect to some target statistics derived from a physical model. In particular, we may penalize the deviation of the height distribution of the water surface with respect to a physically-justified quasi-Gaussian model and drive the surface evolution toward (weakly) satisfying such a global property. This approach was motivated to solve the problem of large kurtosis (spikiness) of the reconstructed surface obtained for some datasets via the base energy model.

5.2.1 Deviation of the height distribution. Cumulative distribution function (CDF) approach.

Let us know introduce some theoretical details. If $Z(u,v) = Z(\mathbf{u})$ is the height of the surface (wave) and it is interpreted as a random variable, then its cumulative distribution function (CDF) can be expressed as

$$\mathrm{cdf}^Z(Z_0) = P(Z \leq Z_0) = \frac{1}{A}\int_U H(Z_0 - Z(\mathbf{u}))\mathrm{d}\mathbf{u},$$

where $H(\cdot)$ is the Heaviside function and $A = \int_U \mathrm{d}\mathbf{u}$ is the area of the (fixed) domain of integration. Suppose (12) is augmented with an extra energy term $\gamma E_{\mathrm{cdf}}(S)$, $\gamma > 0$, that measures the discrepancy in the L^2 sense between a target height CDF that we wish to enforce, $G(Z)$, and the experimental CDF of the height:

$$E_{\mathrm{cdf}}(Z) = \int_{-\infty}^{\infty} \tfrac{1}{2}\big(G(\hat{z}) - \mathrm{cdf}^Z(\hat{z})\big)^2 \mathrm{d}\hat{z}. \tag{79}$$

To compute the first variation of (79), we can directly use the definition of the Gâteaux derivative or augment Z with an artificial time variable, $Z = Z(\mathbf{u}, t)$, so that the energy depends on t, differentiate with respect to this variable and exploit the relationship between both derivatives:

$$\frac{\mathrm{d}}{\mathrm{d}t}E_{\mathrm{cdf}} = \frac{\delta E_{\mathrm{cdf}}}{\delta Z}(Z_t).$$

Carrying out operations in the distributional sense,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}E_{\mathrm{cdf}} &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty}\big(G(\hat{z})-\mathrm{cdf}^{Z}(\hat{z})\big)^2\mathrm{d}\hat{z}\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty}\tfrac{1}{2}\big(G(\hat{z})-\frac{1}{A}\int_{U}H(\hat{z}-Z(\mathbf{u}))\mathrm{d}\mathbf{u}\big)^2\mathrm{d}\hat{z}\\
&= \int_{-\infty}^{\infty}\big(G(\hat{z})-\mathrm{cdf}^{Z}(\hat{z})\big)\big(\frac{1}{A}\int_{U}\delta(\hat{z}-Z(\mathbf{u}))Z_t\mathrm{d}\mathbf{u}\big)\mathrm{d}\hat{z}\\
&= \int_{U}\frac{1}{A}\int_{-\infty}^{\infty}\big(G(\hat{z})-\mathrm{cdf}^{Z}(\hat{z})\big)\,\delta(\hat{z}-Z(\mathbf{u}))\mathrm{d}\hat{z}\,Z_t\mathrm{d}\mathbf{u}\\
&= \int_{U}\frac{1}{A}\big(G(Z)-\mathrm{cdf}^{Z}(Z)\big)\,Z_t\mathrm{d}\mathbf{u}\\
&= \int_{U}\nabla_Z E_{\mathrm{cdf}}\,Z_t\mathrm{d}\mathbf{u},
\end{aligned}$$

where the gradient of (79) with respect to the height is

$$\nabla_Z E_{\mathrm{cdf}}(Z)=\frac{1}{A}\big(G(Z)-\mathrm{cdf}^{Z}(Z)\big), \tag{80}$$

or, in expanded form reflecting the point-wise nature of the gradient,

$$\nabla_Z E_{\mathrm{cdf}}(\mathbf{u})=\frac{1}{A}\Big(G\big(Z(\mathbf{u})\big)-\frac{1}{A}\int_{U}H\big(Z(\mathbf{u})-Z(\mathbf{v})\big)\mathrm{d}\mathbf{v}\Big).$$

As a result of the statistical penalty, a new non-linear term of the form (80) appears in the EL equation (37) (and in (43)), while the boundary condition remains unchanged. It is as if the non-linear term (41) in PDE (37) is replaced by

$$g(Z)\leftarrow g(Z)+\gamma\nabla_Z E_{\mathrm{cdf}}(Z). \tag{81}$$

5.2.2 Probability density function (PDF) approach.

Another reasonable energy to measure the statistical discrepancy between the empirical distribution of the wave field and the one dictated by the physical model is the L^2 difference between probability density functions (PDFs):

$$E_{\mathrm{pdf}}(Z)=\int_{-\infty}^{\infty}\tfrac{1}{2}\big(\dot{G}(\hat{z})-\mathrm{pdf}^{Z}(\hat{z})\big)^2\mathrm{d}\hat{z}, \tag{82}$$

where $\dot{G}(Z)$ is the target PDF that we wish to enforce and

$$\mathrm{pdf}^{Z}(\hat{z})=\frac{\mathrm{d}}{\mathrm{d}\hat{z}}\,\mathrm{cdf}^{Z}(\hat{z})=\frac{1}{A}\int_{U}\frac{\mathrm{d}}{\mathrm{d}\hat{z}}H(\hat{z}-Z(\mathbf{u}))\mathrm{d}\mathbf{u}=\frac{1}{A}\int_{U}\delta(\hat{z}-Z(\mathbf{u}))\mathrm{d}\mathbf{u}$$

is the PDF of the height function Z. Following similar steps as before, the EL equation (37) would have instead an extra term of the form

$$\nabla_Z E_{\mathrm{pdf}}(Z) = -\frac{1}{A}\frac{\mathrm{d}}{\mathrm{d}Z}\big(\dot{G}(Z) - \mathrm{pdf}^Z(Z)\big). \tag{83}$$

Proof. Again, carrying out operations in the distributional sense,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}E_{\mathrm{pdf}} &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty} \tfrac{1}{2}\big(\dot{G}(\hat{z}) - \mathrm{pdf}^Z(\hat{z})\big)^2 \mathrm{d}\hat{z} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty} \tfrac{1}{2}\big(\dot{G}(\hat{z}) - \frac{1}{A}\int_U \delta(\hat{z} - Z(\mathbf{u}))\mathrm{d}\mathbf{u}\big)^2 \mathrm{d}\hat{z} \\
&= \int_{-\infty}^{\infty} \big(\dot{G}(\hat{z}) - \mathrm{pdf}^Z(\hat{z})\big)\big(\frac{1}{A}\int_U \dot{\delta}(\hat{z} - Z(\mathbf{u}))Z_t \mathrm{d}\mathbf{u}\big)\mathrm{d}\hat{z} \\
&= \int_U \frac{1}{A}\int_{-\infty}^{\infty} \big(\dot{G}(\hat{z}) - \mathrm{pdf}^Z(\hat{z})\big)\, \dot{\delta}(\hat{z} - Z(\mathbf{u}))\mathrm{d}\hat{z}\, Z_t \mathrm{d}\mathbf{u} \\
&= \int_U \frac{1}{A}(-1)\frac{\mathrm{d}}{\mathrm{d}z}\big(\dot{G}(Z) - \mathrm{pdf}^Z(Z)\big)\, Z_t \mathrm{d}\mathbf{u} \\
&= \int_U \nabla_Z E_{\mathrm{pdf}}\, Z_t \mathrm{d}\mathbf{u},
\end{aligned}$$

where we used the rule for the n-derivative of the Dirac δ distribution:

$$\int_{x-a}^{x+a} f(t)\delta^{(n)}(x-t)\,\mathrm{d}t = (-1)^n f^{(n)}(x). \qquad \square$$

Therefore, the simplified expression for gradient of the energy with respect to the height $Z = Z(\mathbf{u})$ is (83), which in expanded form is

$$\nabla_Z E_{\mathrm{pdf}}(\mathbf{u}) = -\frac{1}{A}\frac{\mathrm{d}}{\mathrm{d}Z}\Big(\dot{G}\big(Z(\mathbf{u})\big) - \frac{1}{A}\int_U \delta\big(Z(\mathbf{u}) - Z(\mathbf{v})\big)\mathrm{d}\mathbf{v}\Big).$$

5.2.3 Characteristic function approach.

There is yet another approach to enforce some target statistics: enforcing the constraint via the characteristic function of the random variable, i.e., the inverse Fourier transform of the PDF. The gradient of the energy with respect to the height can be computed in the frequency domain and then converted to the spatial domain, yielding the same gradient as the PDF approach.

The characteristic function of the height function is

$$\mathrm{Char}^Z(\omega) = \mathsf{E}[e^{\imath\omega\hat{z}}] = \int_{-\infty}^{\infty} e^{\imath\omega\hat{z}}\, \mathrm{pdf}^Z(\hat{z})\mathrm{d}\hat{z},$$

which can be expressed as an integral in the parameter space U performing calculations in the distributional sense as follows:

$$\begin{aligned}
\mathrm{Char}^Z(\omega) &= \int_{-\infty}^{\infty} e^{\imath\omega\hat{z}} \frac{1}{A} \int_U \delta(\hat{z} - Z(\mathbf{u}))\mathrm{d}\mathbf{u}\mathrm{d}\hat{z} \\
&= \int_{-\infty}^{\infty} \frac{1}{A} \int_U e^{\imath\omega\hat{z}} \delta(\hat{z} - Z(\mathbf{u}))\mathrm{d}\mathbf{u}\mathrm{d}\hat{z} \\
&= \int_{-\infty}^{\infty} \frac{1}{A} \int_U e^{\imath\omega Z(\mathbf{u})} \delta(\hat{z} - Z(\mathbf{u}))\mathrm{d}\mathbf{u}\mathrm{d}\hat{z} \\
&= \frac{1}{A} \int_U \int_{-\infty}^{\infty} e^{\imath\omega Z(\mathbf{u})} \delta(\hat{z} - Z(\mathbf{u}))\mathrm{d}\hat{z}\mathrm{d}\mathbf{u} \\
&= \frac{1}{A} \int_U e^{\imath\omega Z(\mathbf{u})} \int_{-\infty}^{\infty} \delta(\hat{z} - Z(\mathbf{u}))\mathrm{d}\hat{z}\mathrm{d}\mathbf{u} \\
&= \frac{1}{A} \int_U e^{\imath\omega Z(\mathbf{u})} \mathrm{d}\mathbf{u}.
\end{aligned}$$

Define the energy in the frequency domain to measure the discrepancy between a target characteristic function $\Phi(\omega)$ and an experimental characteristic function $\mathrm{Char}^Z(\omega)$:

$$E_{\mathrm{Char}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tfrac{1}{2} |\Phi(\omega) - \mathrm{Char}^Z(\omega)|^2 \mathrm{d}\omega.$$

To compute the first variation of this energy, augment Z with a fictitious time variable, $Z = Z(\mathbf{u}, t)$, so that the energy depends on t, and differentiate with respect to this variable:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} E_{\mathrm{Char}} &= \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Phi(\omega) - \mathrm{Char}^Z(\omega)|^2 \mathrm{d}\omega \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{2\pi} \int_{-\infty}^{\infty} \tfrac{1}{2} \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)^* \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big) \mathrm{d}\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Re\left\{ \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)^* \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)_t \right\} \mathrm{d}\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Re\left\{ \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)^* \big(-\frac{1}{A} \int_U e^{\imath\omega Z(\mathbf{u})} \mathrm{d}\mathbf{u}\big)_t \right\} \mathrm{d}\omega \\
&= -\frac{1}{A} \Re\left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)^* \int_U e^{\imath\omega Z(\mathbf{u})} \imath\omega Z_t \mathrm{d}\mathbf{u}\, \mathrm{d}\omega \right\} \\
&= -\frac{1}{A} \Re\left\{ \int_U \frac{1}{2\pi} \int_{-\infty}^{\infty} \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)^* \imath\omega e^{\imath\omega Z(\mathbf{u})} \mathrm{d}\omega\, Z_t \mathrm{d}\mathbf{u} \right\} \\
&= \int_U -\frac{1}{A} \Re\left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)^* \imath\omega e^{\imath\omega Z(\mathbf{u})} \mathrm{d}\omega \right\} Z_t \mathrm{d}\mathbf{u} \\
&= \int_U \nabla_Z E_{\mathrm{Char}}\, Z_t \mathrm{d}\mathbf{u},
\end{aligned}$$

where $\Re\{s\}$ stands for the real part of $s \in \mathbb{C}$, and

$$\nabla_Z E_{\mathrm{Char}} = -\frac{1}{A} \Re\left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \big(\Phi(\omega) - \mathrm{Char}^Z(\omega)\big)^* \imath\omega e^{\imath\omega Z(\mathbf{u})} \mathrm{d}\omega \right\}.$$

Using the properties of the Fourier transform:

$$\begin{aligned} \mathrm{pdf}^*(z) &\leftrightarrow \big(\mathrm{Char}(-\omega)\big)^*, \\ \frac{\mathrm{d}}{\mathrm{d}z}\,\mathrm{pdf}(z) &\leftrightarrow -\imath\omega\,\mathrm{Char}(\omega), \end{aligned}$$

and the pair of Fourier transforms:

$$\begin{aligned} \mathrm{pdf}(z) &\leftrightarrow \mathrm{Char}(\omega), \\ \dot{G}(z) &\leftrightarrow \Phi(\omega), \end{aligned}$$

we may simplify

$$\begin{aligned} \nabla_Z E_{\mathrm{Char}} &= -\frac{1}{A}\Re\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big(-\imath\omega\big(\Phi(\omega)-\mathrm{Char}^Z(\omega)\big)\Big)^* e^{\imath\omega Z(\mathbf{u})}\mathrm{d}\omega\right\} \\ &= -\frac{1}{A}\Re\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big(-\imath(-\nu)\big(\Phi(-\nu)-\mathrm{Char}^Z(-\nu)\big)\Big)^* e^{-\imath\nu Z(\mathbf{u})}\mathrm{d}\nu\right\} \\ &= -\frac{1}{A}\Re\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big((\imath\nu)\big(\Phi(-\nu)-\mathrm{Char}^Z(-\nu)\big)\Big)^* e^{-\imath\nu Z(\mathbf{u})}\mathrm{d}\nu\right\} \\ &= -\frac{1}{A}\Re\left\{\Big(\frac{\mathrm{d}}{\mathrm{d}z}\big(\dot{G}(z)-\mathrm{pdf}^Z(z)\big)\Big)^*\right\} \\ &= -\frac{1}{A}\Re\left\{\frac{\mathrm{d}}{\mathrm{d}z}\big(\dot{G}(Z)-\mathrm{pdf}^Z(Z)\big)\right\} \\ &= -\frac{1}{A}\frac{\mathrm{d}}{\mathrm{d}z}\big(\dot{G}(Z)-\mathrm{pdf}^Z(Z)\big) \\ &\overset{(83)}{=} \nabla_Z E_{\mathrm{pdf}} \end{aligned}$$

This result agrees with Parseval's theorem, which states that energy can be measured in both spatial and frequency domains:

$$E_{\mathrm{Char}} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tfrac{1}{2}|\Phi(\omega)-\mathrm{Char}(\omega)|^2\mathrm{d}\omega \overset{\text{Parseval}}{=} \int_{-\infty}^{\infty}\tfrac{1}{2}|\dot{G}(\hat{z})-\mathrm{pdf}^Z(\hat{z})|^2\mathrm{d}\hat{z} = E_{\mathrm{pdf}}.$$

Thus, enforcing the statistical constraint via the L^2 difference of characteristic functions is, by Parseval's theorem, equivalent to the previously described PDF approach.

5.2.4 Theoretical models of wave height distributions

Theoretical probabilistic wave models that can be used as target physical distributions in the statistical regularizers are presented in the works of Tayfun and Fedele [98, 22]. These models are quasi-Gaussian distributions that capture the asymmetry present in real life

water waves, which have steep crests and shallow troughs [56]. The Gram-Charlier (GC) model is based on the approximate PDF:

$$p_Z(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\big(1 + \frac{\lambda_3}{6}x(x^2-3) + \frac{\lambda_3^2}{72}(x^6 - 15x^4 + 45x^2 - 15) + \frac{\lambda_4}{24}(x^4 - 6x^2 + 3)\big), \quad (84)$$

where $x = (z-\mu)/\sigma$ is the normalized height z in terms of μ and σ, which are the mean and standard deviation of the height distribution. The parameters λ_3 and λ_4 are the skewness and kurtosis (third- and fourth-order moments) of the wave height distribution, respectively. From (84), different models may be considered depending on whether all or a subset of the terms of the polynomial are used in the model. Observe that the GC models consist of a Gaussian distribution perturbed by a low-order polynomial (on the skewness and kurtosis) that characterizes the non-linear features of ocean waves. In practice, parameters μ, σ, λ_3 and λ_4 can be estimated from empirical data by their sample estimators. This model was used for analyzing reconstructed surfaces by the variational method of section 2.3 and was reported in [29]. Other models of wave height distributions are also derived by Juglard et al. [94].

5.2.5 Numerical considerations

According to (81), a numerical implementation of the previously presented statistical regularizers is possible in the same multigrid framework of chapter 3 for the basic variational graph method. A von Neumann analysis can be applied to derive an approximate ℓ^2 stability condition for the new time-stepping smoothers of the height function in the multigrid method. To compute the non-linear term (81) due to the statistical regularizer, an estimation of the empirical wave height distribution (CDF or PDF) is needed. The histogram of the height function at the finest level of the V-cycles is used to compute an approximation of the empirical wave height distribution that is used throughout the V-cycle. This implies a classification process of the range of the current wave height function into bins. Numerically, the new non-linearity requires an additional check on the time step: the maximum height update must be of the order of the bin size for the discrete approximation of the empirical wave height to be valid. Depending on the bin size or the number of bins used for a fixed range of the height function, the new constraint may require smaller time steps than

those given by the ℓ^2 stability analysis. Hence, more iterations may be needed to achieve convergence, thus the multigrid solver may be slower than that which does not incorporate the statistical regularizer. We refer the reader to the upcoming chapter on experiments to show the effect of the statistical regularizer on the shape of the reconstructed surface.

5.3 Modification of the photometric criterion

The variational reconstruction model presented so far is a sensible model, but there are still some situations where its performance may deteriorate. For example, the model is designed for scenes with Lambertian objects (i.e., whose surface obey the Lambertian light reflection hypothesis), but in practice, images of ocean waves acquired under favorable circumstances only follow this rule approximately and therefore, the reconstruction of the waves worsens as the reflectance of the scene departs from the Lambertian case. In addition, breaking ocean waves generate localized white regions on the water surface that contrast with the darker neighboring intensities. According to the proposed data fidelity energy (13)-(15), a mismatch in the reconstruction of these regions contributes more to the total energy (in a quadratic manner) than mismatched regions with smaller intensity jumps. Therefore, to minimize the composite energy, the descent flow that drives the surface shape evolution concentrates its efforts in reducing the error in these regions, generating localized high frequencies that are not present in the physical surface of the waves. Thus, in these regions, the model generates undesired artifacts that must be corrected. This has been observed and confirmed through experimental tests with real data. In summary, there is some room for improvement in the existing model and two approaches have been investigated to address both aforementioned conflictive situations that may cause a breakdown of the reconstruction given by the proposed model in chapter 3.

5.3.1 Quadratic vs. Sub-quadratic error functions

One approach developed to improve the performance of the base model in the aforementioned situations is to slightly modify the photometric criterion within the data-fidelity energy. So far, a quadratic (point-wise) law (15) has been investigated to measure the photometric error between observed image data and the radiance model f. Now, instead we

investigate functions that grow sub-quadratically so that photometric errors are weighted differently according to their value, and large errors do not capture the attention of the gradient descent flow. This same idea of weighting the residuals differently across the scale is closely related to the theory of robust M-estimators in Statistics (see [113] and the appendix in [35]). There are many functions that can be tested, but all of them will increase the amount of non-linearity in the model and in its necessary optimality conditions (PDEs). The advantage of the quadratic base model (15) is that it produces a linear radiance PDE (39), linear equations for the AGC coefficients (56), and mild non-linearities in the height PDE (41).

An alternative solution to the aforementioned problem could be achieved by incorporating intensity dependence in the regularizers, e..g., $E_{\text{geom}} = \int_U \frac{1}{2}\|\nabla Z\|^2 \rho(\|\nabla I_i\|)\mathrm{d}\mathbf{u}$, where ρ is some monotonically increasing function that raises smoothness costs at high intensity gradients. This coupling idea discourages discontinuities in the unknowns to coincide with image intensity edges. In its opposite version (ρ monotonically decreasing instead of increasing) this idea accounts for the good performance of other energy optimization approaches designed to transfer image discontinuities to the unknowns of the problem [3, 30, 26]. This approach modifies the diffusive terms in the PDEs and has not been explored in this thesis, but it could be interesting to test in future research.

Assuming that a generic photometric criterion $\phi_i \equiv \phi(r_i)$, with $r_i = I_i - f$, is used in (14), how do the necessary optimality PDEs change to account for it? The derivative of ϕ, denoted by $\dot{\phi}$, plays a prominent role in the answer to this question. This fact is not surprising at all since the derivative of the quadratic function in the base model is the source of linearity of the non-principal part of the PDE in the radiance and the linearity in the solution of the AGC coefficients. The necessary optimality PDEs reflect the choice of a generic photometric criterion ϕ as follows. The new radiance PDE is

$$\beta\Delta f - \sum_{i=1}^{N_c} \dot{\phi}_i(I_i - f)\, \mathtt{J}_i = 0. \tag{85}$$

If $\beta = 0$, the optimal solution of (85) satisfies

$$\sum_{i=1}^{N_c} \dot{\phi}_i(I_i - f)\, \mathtt{J}_i = 0,$$

which only yields a simple solution in the linear case (quadratic ϕ). The correction of the data fidelity term does not affect the boundary condition (40).

It is remarkable that the new height PDE still does not depend on ϕ, but solely on its derivative: in the derivation of the optimality condition, the term in ϕ cancels out. The formula for the flow in divergence form (157) is the shortest path to obtain the desired answer (241):

$$\nabla_S E_i = |\mathtt{M}^i| \tilde{Z}_i^{-3} \, \dot{\phi}_i (I_i - f) \, (\mathbf{X} - \mathbf{C}_i) \cdot \nabla \hat{f}$$

Therefore, the necessary optimality condition becomes (37), but with the non-linear term (41) replaced by

$$g(Z, f) = \sum_{i=1}^{N_c} |\mathtt{M}^i| \tilde{Z}_i^{-3} \, \dot{\phi}_i (I_i - f) \, (u - C_i^1, v - C_i^2) \cdot \nabla f.$$

Consistency with known results is maintained, e.g., particularizing the previous expression for the quadratic ϕ gives (41). The homogeneous Neumann boundary condition that accompanies the PDE remains unchanged using the same argument as in section 3.3.

There are many possible choices of ϕ that one can test to improve the results of the basic model in the conflictive non-Lambertian regions. It would take a lot of time to test different types of error functions by reconstructing a long sequence of stereo images and compare the results to choose the best function among them. Ultimately, ϕ is, as long as it is supported by experimental results, a design choice like many other parameters: the number of pre- and post-relaxation sweeps at each level of the multigrid method, the number of V-cycles, the processing schedule to solve the PDEs, the value of the weights α, β, etc. In this thesis, the following sub-quadratic function, displayed in Figure 22, is proposed instead of the more immediate choice $\phi(r) = |r|$, which is non-differentiable (or has a discontinuous generalized derivative) at $r = 0$:

$$\phi(r) = \frac{r^2}{1 + (r/a)^2}, \tag{86}$$

$$\dot{\phi}(r) = \frac{2r}{\left(1 + (r/a)^2\right)^2}, \tag{87}$$

$$\ddot{\phi}(r) = \frac{2(1 - 3(r/a)^2)}{\left(1 + (r/a)^2\right)^3}, \tag{88}$$

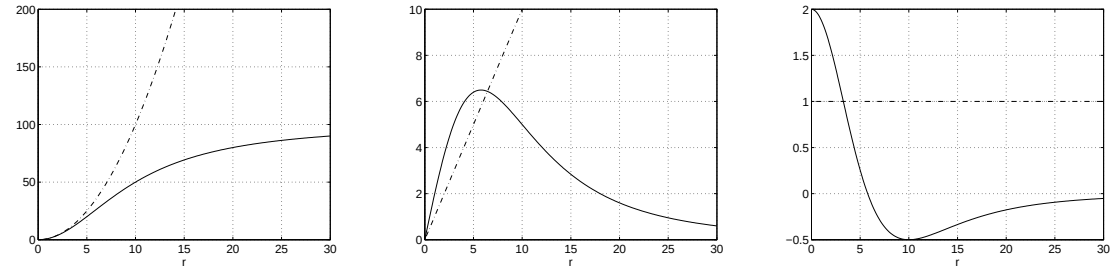


Figure 22: Comparison of sub-quadratic photometric error function (solid line) $\phi = r^2/(1+(r/a)^2)$, with $a = 10$, and quadratic error function $\phi = \frac{1}{2}r^2$ (dashed line). Left: error functions, $\phi(r)$. Center: first order derivatives, $\dot{\phi}(r)$. Right: second order derivatives, $\ddot{\phi}(r)$.

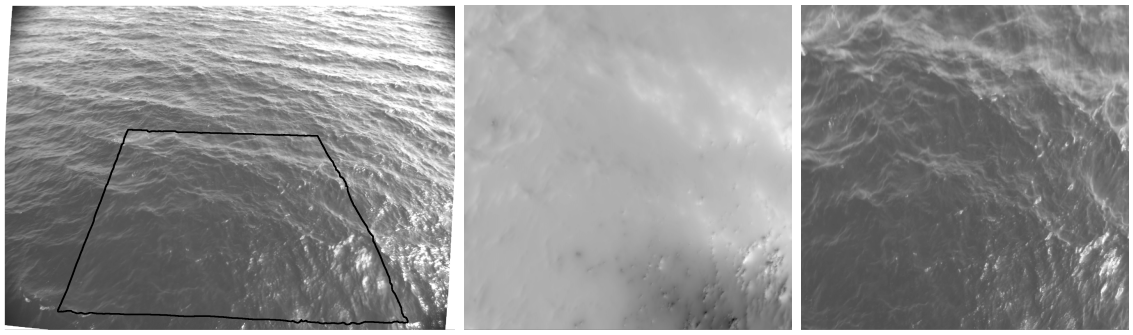


Figure 23: Quadratic photometric criterion, (15). Left: Modeled image (highlighted region) superimposed on original image. Center and right: Height (grayscale encoded) and radiance functions.

where $a > 0$ is a parameter that allows the designer to control the shape of the function (the value of the intensity error r at which saturation of the penalty begins). For small values of r, both quadratic and sub-quadratic functions behave similarly, but as r increases, the proposed function (86) does not grow as fast as the quadratic penalty, exhibiting a saturation effect. For large intensity errors, the penalty (86) is approximately constant: $\phi(r \gg a) \approx a^2$.

In practice, to simplify the implementation and avoid slowing down the solution of the reconstruction problem, only the height PDE incorporates the previous modification of the photometric criterion: the radiance PDE is unchanged. After all, the radiance is just an auxiliary variable to increase the basin of attraction of the shape minimizer and to avoid image derivatives.

Results obtained with different photometric criteria are reported in Figures 23 and 24. The reflection of the sunlight on the water surface generates highlights or specularities that

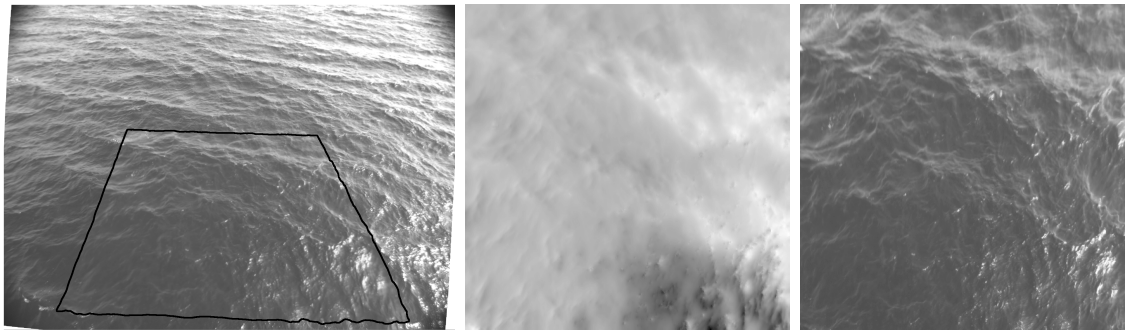


Figure 24: Sub-quadratic photometric criterion. Left: Modeled image (highlighted region) superimposed on original image. Center and right: Height (grayscale encoded) and radiance functions.

appear around the bottom right corner of the images. With the quadratic error function, the multigrid optimization algorithm does not provide a smooth surface: the evolving height function undergoes wild variations in these non-Lambertian regions. Even in case of using a relatively large value of α (weight of the shape regularizer) that keeps the shape of the Lambertian regions smooth, the specular regions can have a violent height variation. Decreasing α to capture shape details in the Lambertian regions causes the specular regions to drive most of the attention of the optimization algorithm since the photometric error concentrates in these regions. The sub-quadratic penalty (86) ameliorates this behaviour. In the modeled images of Figures 23 and 24 is difficult to notice the difference between the results of both quadratic and sub-quadratic functions. However, the difference clearly stands out if the heights of the estimated surface are compared. Using the quadratic error function, the height of the Lambertian areas appear smooth and flat compared to the regions with specularities, where wild variations of the height function occur. In contrast, utilizing the sub-quadratic error function, the height in the specular regions does not have a wild variation and the details in the Lambertian regions are distinguished better.

The second derivative of ϕ, denoted by $\ddot{\phi}$, arises in $\dot{g}(Z)$, which is used in the approximate ℓ^2 stability condition for the time-stepping height smoother, as shown in (246), and in the linearization (49) of the height PDE.

5.3.2 Image pre-processing followed by quadratic error function

A different approach to attenuate the effect of specularities and sea foam on the reconstruction algorithm consists of using the original variational graph method (quadratic error function) but on a transformed version of the data images. A point-wise logarithmic pre-processing of the images has the effect of stretching dark grayscale intensity levels and compressing white intensity levels. The latter is of most significance to our ocean wave application because it compresses large intensity jumps near white regions caused by specularities and sea foam. For images with intensity levels in the range $[0, 255]$, the transformation rule is given by

$$y(x) = \frac{255}{\log(1 + 255c)} \log\big(1 + cx\big), \tag{89}$$

where x and y are the input and output intensities, respectively, both in the range $[0, 255]$. The parameter $c > 0$ controls the shape of the intensity transformation law, i.e., the amount of stretching and compression for dark and white intensity levels, respectively. In practice, it has been observed that the simple choice $c = 1$ yields good results for images like those in Figure 24.

Thus, the pre-processed version of the images can be fed to the variational graph method with quadratic photometric criterion to yield the height and radiance that minimize the composite energy functional on the pre-processed images. Theoretically, the pre-processing of the image intensities does not affect the minimizer for the height function; it only affects the minimizer for the radiance. The true shape of the water surface is the same in both, original and pre-processed images, because we are only changing the "color" sitting on top of the surface. In practice, it does affect because the time steps taken in the gradient descent algorithm significantly increase the convergence rate of the iterative solver, which is a positive effect. This becomes clear by recalling the role that image and radiance derivatives play in the formula for the maximum time step allowed to evolve the height PDE, (54): $\Delta t \propto (\max |\dot{g}(\bar{Z}_{i,j})|)^{-1}$, where the analytical formula for $\dot{g}(Z)$ is given by (247). The logarithmic transformation decreases the value of $\max |\dot{g}(\bar{Z}_{i,j})|$, therefore increasing the time step and the convergence rate of the non-linear height PDE.

Observe, however that the radiance significantly changes (still via a linear PDE) to adapt to the logarithmic input images. The relationship between the weights of the different terms in the composite energy functional is also affected by the logarithmic transformation. A larger β is required to yield the same amount of smoothness as in the case of original images. The composition of the logarithmic image transformation and the quadratic photometric criterion results in an overall photometric error function (on the original intensities) that is similar to the sub-quadratic penalty discussed in the previous section.

Finally, the recovery of the radiance function for the original, unprocessed images can be achieved by applying the inverse (exponential) transformation to the radiance function returned by the variational graph method acting on the logarithmic images:

$$x(y) = \frac{1}{c}\left(\exp\Big(y\frac{\log(1+255c)}{255}\Big) - 1\right). \tag{90}$$

5.4 Conclusion

In this chapter, several improvements of the base variational graph reconstruction method of chapter 3 have been investigated. On the one hand, some initial modifications of the regularizers have been developed within our variational framework to take into account physical properties of ocean waves during the reconstruction process. These properties would be difficult to take into account in other stereo methods, but in our case they can be easily handled as a consequence of our method being founded upon a physical model of the scene. The geometric area-based regularizer affects (i) the diffusive term of the height PDE, and by extension, (ii) the inner product used to measure the data fidelity term. The statistical regularizers allows us to weakly enforce global wave height distribution models on the reconstructed surface, and they appear in the height PDE as new non-linear terms. On the other hand, an improvement and generalization of the photometric matching criterion has been introduced to overcome the weaknesses of the Lambertian generative model of the images in presence of specularities and/or abrupt intensity changes caused by sea foam. In particular, the choice of a sub-quadratic error function or a logarithmic image pre-processing have been analyzed. Multigrid is still a valid numerical framework to implement the aforementioned modifications of the variational graph method.

CHAPTER VI

EXPERIMENTS WITH SEQUENCES

The variational reconstruction method designed for still images in chapter 3 can be naturally extended to process multiple temporal images, that is, stereo video. The simplest way to do so is to process images sequentially, on a snapshot by snapshot basis, via estimation of the new surface shape and radiance based on the previously reconstructed surface. In this way, the redundancy in the slow varying nature of ocean waves, compared to the video frame rate, is exploited to avoid computing the solutions of each snapshot's optimality PDEs from scratch. This is the processing scheme used in this chapter to apply the variational graph method to empirical video data collected in field experiments.

Image-based photometric scores are used to measure the quality of the reconstructed surface according to the modeled images that the variational method generates from the estimated shape and radiance of the ocean waves. The most immediate score is the data-fidelity term of the composite energy functional. In addition, since there is no available ground truth to compare the estimated surface in 3-space, a validation based on empirically-based physical models of ocean waves is also required.

6.1 Comparison of the effects of the regularizers

After validating the numerical implementation of the proposed variational stereo method with synthetic data, some experiments with real data are carried out. The input images from a field experiment at the San Diego Pier are of size 1008×1008 pixels. This data was collected by Dr. Benetazzo [4]. At the finest level of the multigrid algorithm, the gradient descent PDEs are discretized on a 2-D grid with 513×513 points. The distance between grid points is $h = 1.5$ cm. Therefore, the grid covers an area of approximately $7.7 \times 7.7\, m^2$. Recalling the comments on the examples given in section 3.5, the distance between adjacent grid points h is chosen to match the image resolution. We test the method on a sequence of 10 consecutive images and show averaged results. The deforming surface is initialized by

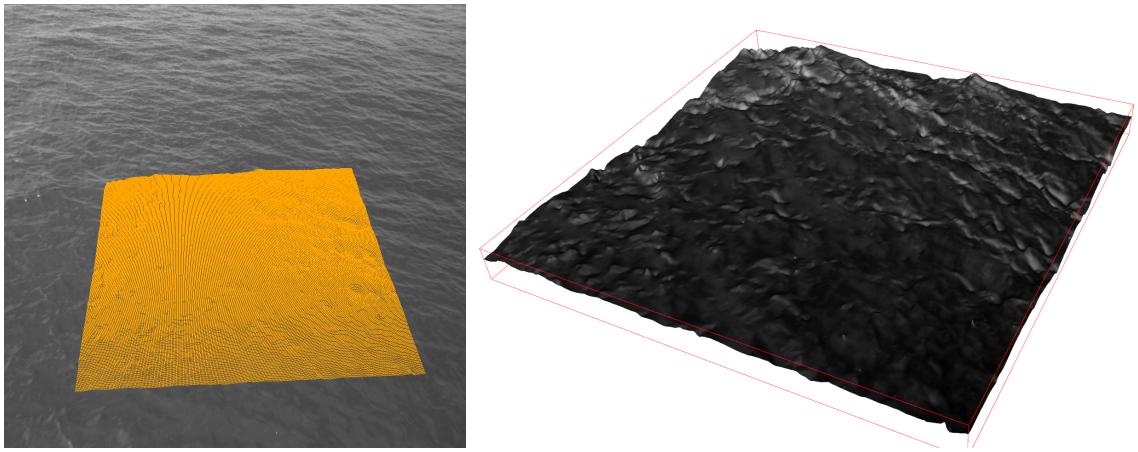


Figure 25: Left: projection of the estimated water surface on image 2. The graph or elevation map has been obtained on a grid of 513×513 points. Right: Texture-mapped surface: perspective, three-dimensional representation of the estimated model of the water surface consisting of the shape (height) and radiance of the surface.

the plane $Z = 0$. A multigrid method (Figure 9) with 6 levels and 200 V-cycles (with 1 pre- and post-relaxation sweeps per level) is used to solve the problem at each snapshot. For the first snapshot, a full multigrid method (FMG) with 200 V-cycles per level is performed prior to entering the above processing schedule.

First we report the results of the variational graph method described in chapter 3. Figure 25 shows the surface obtained upon convergence for one of the snapshots. The corresponding height and radiance functions are displayed on Figure 26. Combining both functions, a textured 3-D plot of the surface is generated, as shown in Figure 25. The weights of the regularizers are empirically determined: $\alpha = 4 \cdot 10^{-2}, \beta = 4 \cdot 10^{-3}$ (and $\gamma = 0$). The first row of Table 2 shows averaged quantitative results for this scenario. The third column of Table 2, $\tilde{E}_{\text{data}}$, represents the photometric consistency per cell:

$$\tilde{E}_{\text{data}} = \frac{E_{\text{data}}}{N} \approx \frac{h^2}{N} \sum_{i=1}^{N_c} \sum_{j=1}^{N} \phi_i(\mathbf{u}_j) \mathtt{J}_i(\mathbf{u}_j), \tag{91}$$

where the number of cells that tessellate the parameter space is $N = 513^2$ and the number of cameras is $N_c = 2$. $\tilde{E}$ and $\tilde{E}_{\text{data}}$ have units of (intensity levels)^2(pixels)^2/cell; as is customary, images and radiance are in the range of [0,255] intensity levels. Being errors equally distributed in both images and $\text{Area}(\Omega_i) \approx 0.29\text{Area}(I_i)$, a value $\tilde{E}_{\text{data}} = 7.75$ implies an RMS error of $\sqrt{2E_i/\text{Area}(\Omega_i)} \approx 2.63$ intensity levels, which is a good fit for

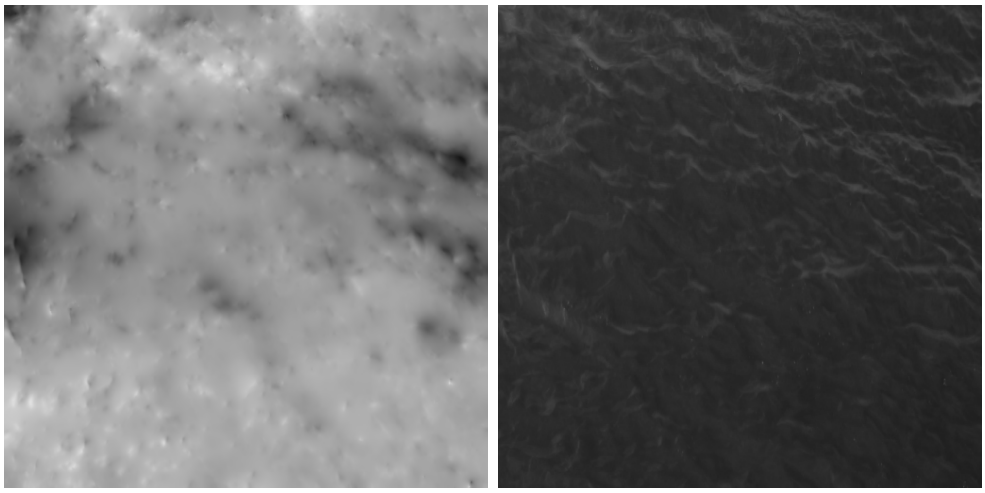


Figure 26: Left: estimated height function $Z(u, v)$ (shape of the water surface). Height is represented by greyscale intensities, from dark (low) to white (high). Right: estimated radiance function $f(u, v)$ (texture on the surface).

Table 2: Energy and statistics of the reconstructed sequence

Exp #	$\tilde{E}$	$\tilde{E}_{\text{data}}$	S/U	Z_{pp}	σ	skewness	kurtosis
1	7.90	7.74	1.0154	31.36	3.39	-0.95	2.05
2	7.93	7.76	1.0163	31.60	2.94	-1.10	2.77
3	7.90	7.75	1.0140	30.01	3.35	-0.19	0.52
4	7.94	7.78	1.0148	28.77	2.94	-0.21	0.66

$|I_i - f|$ in a [0,255] scale. The ratio S/U is the area of the surface divided by the area of the domain plane. Z_{pp} is the peak-to-peak wave height $Z_{pp} = Z_{\max} - Z_{\min}$, and σ is the wave height standard deviation, both are measured in cm. The skewness and excess kurtosis appear in the last columns of the table.

In the second row of Table 2, we show the results of the variational reconstruction method with the area-based regularizer (74) and the same weights as in experiment 1. Quantities are very similar to those of experiment 1, but the resulting surface has on average a smaller standard deviation and a larger kurtosis.

In experiment 3 (third row of Table 2), the statistical CDF regularizer with $\gamma/A = 10^{-2}$ and Gaussian distribution is added to the variational reconstruction method of experiment 1. Experiment 4 is similar to experiment 3, but with the area-based regularizer (74). By comparing rows 1 vs. 3 and 2 vs. 4 we observe that the statistical regularizer has a

direct impact on the skewness and kurtosis of the surface while the rest of the measurements remain almost unchanged. Therefore, we are able to (weakly) enforce a physical model of the statistics of the waves while still fitting the observed image measurements. The statistical penalty barely affects the omni-directional spectrum since the latter mostly depends on α, which is the same for all experiments.

In summary, the results of the mean curvature diffusive term are a minor modification of the ones obtained with the Laplacian diffusive term, whereas the statistical regularizer has a stronger impact in the shape of the surface while keeping the photometric fit almost unaltered.

6.2 Experiments at the Black Sea

In other experiments, Dr. Benetazzo collected data in an experimental field setup at one of the MHI NASU (Marine Hydrophysical Institute National Academy of Sciences of Ukraine) platforms. The oceanographic platform is located near the southern seashore of the Crimean peninsula, in the Black Sea. The tower is equipped with meteo-oceanographic instrumentation to monitor wind speed, temperature, humidity, rain, solar radiation, directional waves and tides.

6.2.1 Stereo image acquisition setup

The WASS setup installed at the platform consists of a pair of monochrome progressive cameras mounted on the top floor of the platform, 12 meters above the sea level (subject to tidal fluctuation) and 2.5 meters apart. The cameras operate at a resolution of 2 Megapixels (1624×1236 pixels) and a frame rate (snapshot rate) of 10 Hz. They provide an 8 bit output (256 grayscale intensity levels), and are paired with 6-mm wide angle focal length low distortion lenses. The cameras are jointly triggered and synchronized with a maximum delay of 1 ms using an external electronic board. Cameras are pre-calibrated by means of the publicly available Camera Calibration Toolbox [6]. An example of the stereo images acquired by this setup is given in Figure 17.

6.2.2 Sequential reconstruction using VWASS

The sequential variational graph method was used to reconstruct the surface of the ocean waves. A sequence of 2000 stereo images (snapshots) of size 1624×1236 pixels acquired at 10 Hz was reconstructed using a 6-level multigrid method (Figure 9) with 500 iterations per snapshot, $NV_Z = NV_f = 2$ V-cycles per iteration and one pre- and post-relaxation sweeps per level. The continuous problem was discretized on a grid of size 513×513 with a spatial resolution of $h = 2.5$ cm, thus covering an area of approximately $12.8 \times 12.8\, m^2$. Figure 24 shows the approximate region of interest occupied by the projection of the reconstructed surface on one of the images. This region is chosen to correspond to the near field of the cameras to have more accurate measurements than those obtained if the region corresponds to the far field of the cameras. Figure 24 also shows an example of the type of output (height and radiance functions) given by the variational graph method for this sequence of stereo images. The surface corresponding to the first stereo pair was initialized by zero height plane, i.e., the domain plane of the graph, and reconstructed using a full multigrid method with 500 iterations per level. The weights of the regularizers were chosen empirically: $\alpha = 0.1$ and $\beta = 0.025$. With the above multigrid parameters, the reconstruction of each stereo pair takes approximately 2 minutes on a desktop computer with an Intel Core2Duo Central Processing Unit (CPU) operating at 2.66 GHz.

6.2.3 Image-based validation

To validate the reconstruction from an image-processing point of view, we present the magnitude of the data fidelity term of the composite energy functional proposed in chapter 3 at the end of the 500 iterations performed for each snapshot. Figure 27 shows the RMS value of the photometric score $E_i/\text{Area}(\Omega_i)$ averaged over both images. That is, if the reconstructed surface projects on a fraction of the image area, $\text{Area}(\Omega_i) = r\text{Area}(I_i)$, the aforementioned score is, in terms of (91),

$$\text{RMS}\left[\frac{E_i}{\text{Area}(\Omega_i)}\right] \approx \left(\frac{\tilde{E}_{\text{data}}}{2}\frac{N}{\text{Area}(\Omega_i)}\right)^{\frac{1}{2}} = \left(\frac{N}{2r\text{Area}(I_i)}\right)^{\frac{1}{2}}\sqrt{\tilde{E}_{\text{data}}}. \tag{92}$$

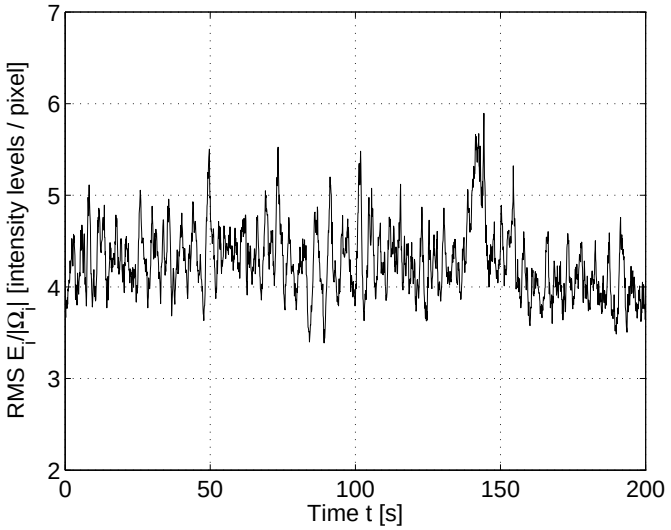


Figure 27: Approximate value of the density photometric score $E_i/\text{Area}(\Omega_i)$, averaged over both cameras, for all the snapshots in the sequence.

Table 3: Comparison of photometric error (92) for two methods used to reconstruct the same stereo image sequence: the variational graph method (chapter 3) and the plane-fitting method (induced homography method) of chapter 4.

Method	mean	standard deviation
Graph	4.27	0.37
Plane-fitting	5.07	0.60

The projected surface gives $r \approx 0.31$, resulting in a conversion factor $\left(\frac{N}{2r\text{Area}(I_i)}\right)^{\frac{1}{2}} = 0.46$. Observe that the RMS values in Figure 27 are intensity levels on a [0,255] scale. The variational graph method provides modeled images with a very good fit to the original ones.

In Table 3, we compare the photometric score of Figure 27 to that given by the flat surface computed by the method of chapter 4 (fitting a plane through the scene). In the second method, the radiance on the surface was computed using (188) (as if $\beta = 0$ in the energy functional). Even though this is not strictly a fair comparison because in the second method the plane is allowed to change from snapshot to snapshot, the energy of the latter is still higher than that of the variational graph method.

6.2.3.1 *Photometric comparison by transferring intensities through the disparity map*

In this section, we compare the variational graph method to the variational dense depth disparity map method (or, simply, *variational disparity method*) of Alvarez et al. [3] that minimizes

$$E(\lambda) = E'_{\text{data}}(\lambda) + \tilde{\alpha} E_{\text{smooth}}(\lambda) \tag{93}$$

with respect to the tangential disparity λ: the component of the disparity map along epipolar lines. For a point correspondence, $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2 = \mathbf{x}_1 + \mathbf{d}$, the disparity is $\mathbf{d} \equiv \mathbf{x}_2 - \mathbf{x}_1 = \gamma\mathbf{N} + \lambda\mathbf{T}$, where $\mathbf{T}$ and $\mathbf{N}$ are the unit tangent and normal vectors of the epipolar line through $\mathbf{x}_2$, respectively. In the case of rectified images, the Euclidean distance from $\mathbf{x}_1$ to the epipolar line through $\mathbf{x}_2$ is $\gamma = 0$ and, therefore, $\|\mathbf{d}\| = \lambda$. A prime notation $'$ is used to distinguish the data fidelity term in (93) from that of the energy minimized by the variational graph method. Since our goal is not to preserve image discontinuities in the disparity map [3] due to the fact that the surface of the ocean is smooth, we alter the smoothness regularizer and choose a non-image dependent penalty:

$$E_{\text{smooth}}(\lambda) = \frac{1}{\text{Area}(\Omega)} \int_{\Omega} \tfrac{1}{2} \|\nabla\lambda\|^2 \, \mathrm{d}\mathbf{x}_1,$$

where $\mathbf{x}_1 \in \Omega$ lies in the reference image (origin for the disparity map). The photometric mismatch between the stereo images for the estimated disparity map is given by the data fidelity term:

$$E'_{\text{data}}(\lambda) = \frac{1}{\text{Area}(\Omega)} \int_{\Omega} \tfrac{1}{2} \big(I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)\big)^2 \, \mathrm{d}\mathbf{x}_1, \tag{94}$$

where $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2(\lambda)$ are corresponding points in images 1 and 2, respectively, with observed intensities I_1 and I_2. This data fidelity energy is not symmetric with respect to the role of each image within a snapshot, but it still serves as a means to compare the errors given by both aforementioned variational methods.

The reconstruction of the sequence was carried out by extending the still-image variational disparity method in the same way that the variational graph method was extended: snapshot by snapshot and using the current disparity to initialize the disparity for the next snapshot. A smoothness weight of $\tilde{\alpha} = 400$ was empirically chosen to provide a smooth

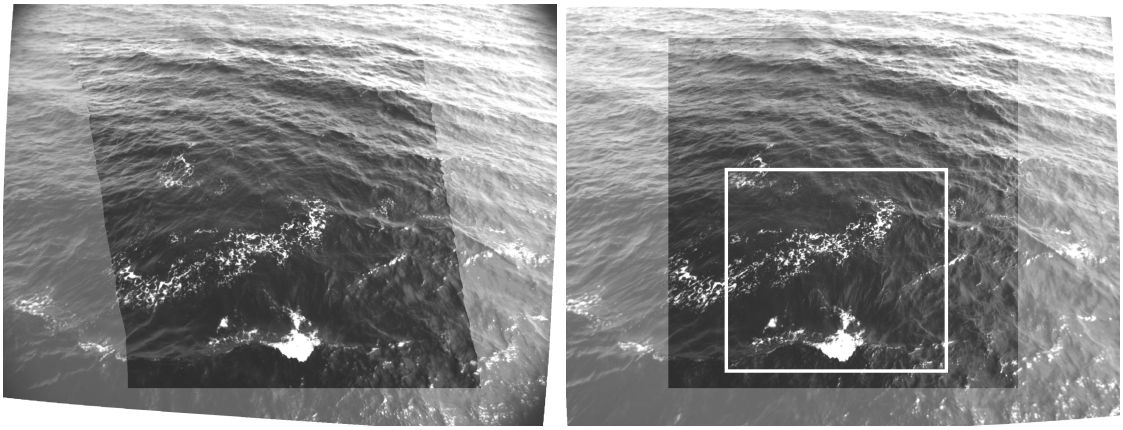


Figure 28: Variational disparity method. Modeled images by symmetric transfer of intensities according to the correspondence given by the disparity map. The right image also shows the region of interest (white rectangle) for numerical comparison in Table 4.

disparity map. The PDE solver for the EL equation of (93) is implemented using multigrid methods, instead of the scale-space approach originally proposed in [3]. Figure 28 shows an example of the modeled images obtained by symmetric transfer of intensities (borrowed from the original images) according to the correspondence given by the disparity map. More specifically, the reconstruction of the surface using the variational disparity method was carried out using an 8-level multigrid method with 200 iterations per snapshot, $NV_Z = NV_f = 2$ V-cycles per iteration and one pre- and post-relaxation sweeps per level. The rectangular integration region Ω in the reference image (right image) was discretized on a grid of size 1025×1025 pixels, as shown in the right image of Figure 28.

Figure 29 shows the same snapshot as in Figure 28, but reconstructed using the variational graph method. Notice the conceptual difference between both methods, image-based vs. object based: in the disparity method, the integral in the data-fidelity energy is over a region (a rectangle) in the image, whereas in the graph method, the integral is over the parameter space, a rectangle in 3-D space (that projects onto a trapeziodal shape in the images). Table 4 summarizes the numerical photometric comparison of both methods in an overlapping region Ω of size 655×600 pixels in the reference image (right image). The disparity map for the graph method was computed from the correspondence induced by the shape of the surface, $Z(u, v)$.

The data fidelity column in Table 4 is proportional to the difference (per pixel) between

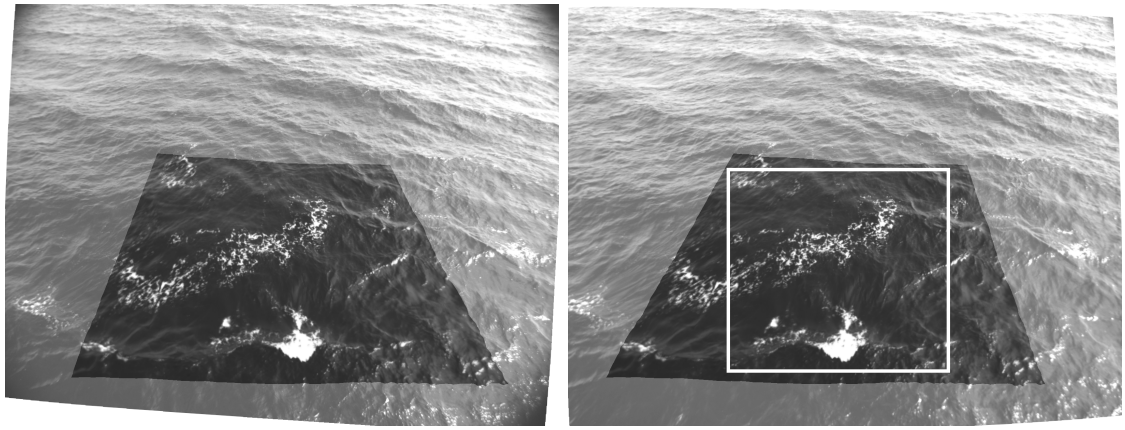


Figure 29: Variational graph method. Modeled images superimposed on the original images. The right image also shows the region of interest (white rectangle) for numerical comparison in Table 4.

Table 4: Comparison of photometric error (94) for two variational methods used to reconstruct the same stereo image sequence.

Method	$(E'_{\text{data}}/2)^{\frac{1}{2}}$		$(\tilde{\alpha}E_{\text{geo}}(\lambda))^{\frac{1}{2}}$	
	mean	standard deviation	mean	standard deviation
Dense disparity	4.43	0.46	2.01	0.14
Graph	5.08	0.54	1.78	0.28

image intensity and radiance function

$$(\tfrac{1}{2}E'_{\text{data}})^{\frac{1}{2}} \propto \tfrac{1}{2}\sqrt{(I_1 - I_2)^2} = \tfrac{1}{2}\sqrt{(I_1 - f + f - I_2)^2} = |I_1 - f|,$$

assuming that, in the variational disparity method, we choose the radiance model to consist of the average of the intensities at corresponding points, $f = (I_1 + I_2)/2$. This way, the values in Tables 3 and 4 are roughly comparable despite the differences in the details of their computation. Both variational methods provide a good fit to the image data. The photometric score used in Table 4 favors the variational disparity method because it is the same error measure that the disparity method is designed to minimize. Nevertheless, Table 4 also shows that the disparity map corresponding to the graph method is smoother than that of the disparity method. Therefore, one could argue that, by balancing smoothness and data-fidelity energies, the solutions given by both methods would be closer photometric-wise.

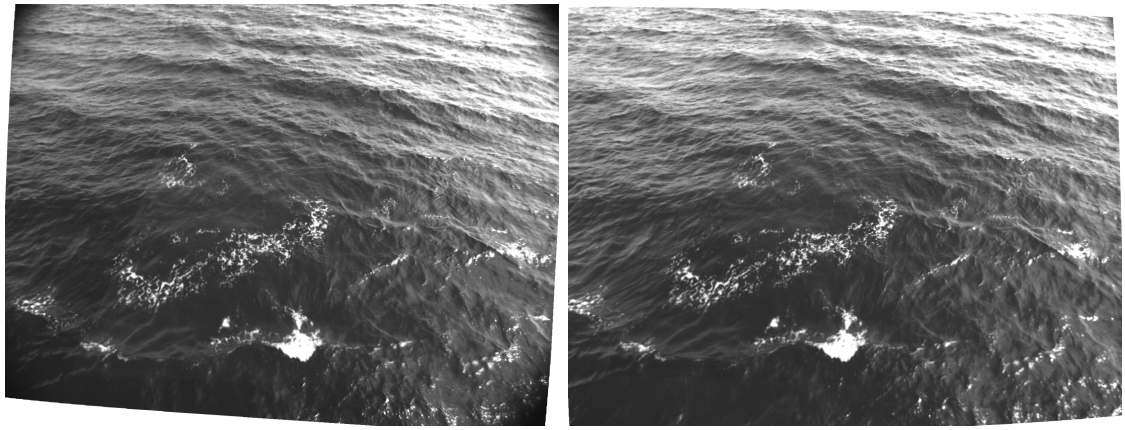


Figure 30: Variational graph method. Modeled images superimposed on the original images.

For completeness, Figure 30 shows, for the same snapshot as in Figure 29, the modeled images superimposed on the original images without intensity transformation outside the region covered by the projection of the candidate surface. Notice that, from an image processing point of view, the reconstructed surface shape and radiance produce modeled images that accurately match the original images, both subjectively according to the perception of the reader's human visual system and objectively, by quantitative measures previously discussed in Figure 27 and Table 4. Figure 31 shows the magnitude of the error between the observed (original) images and the modeled images by the reconstruction method. Observe that large photometric errors occur at large intensity jumps caused by sea foam and/or specularity effects, whereas small errors take place in regions with smooth intensities.

6.2.4 Wave spectra

The four-dimensional reconstructed wave surface can be represented in the form of a space-time volume of wave heights, $V = Z(x, y, t)$, as visualized in Figure 32, where the oscillating pattern of the waves is evident by the oscillating color patterns. The spectra and statistics of the waves can be computed from the reconstructed surface. To capture short waves, some post-processing may be required [90]. High-pass filtering in space is essential to remove the mean surface that masks the superimposed short waves, i.e., to extract useful information on

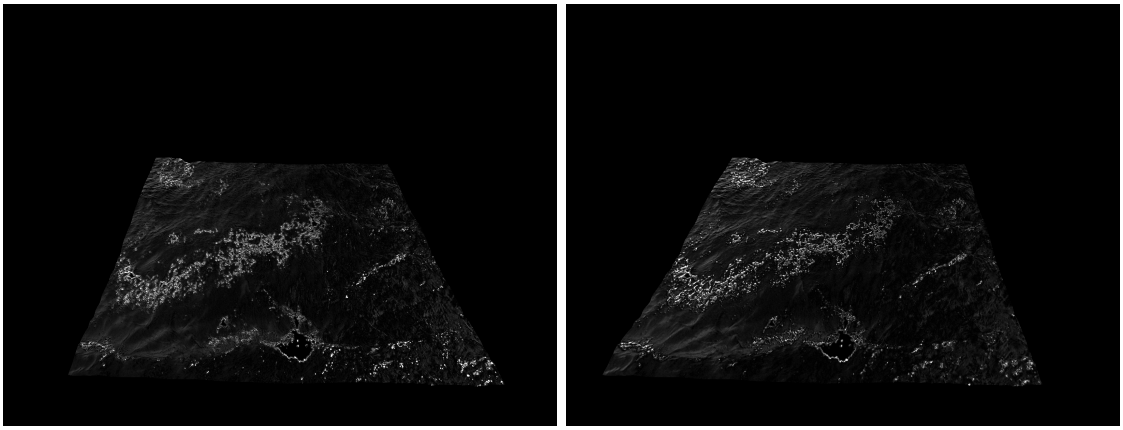


Figure 31: Variational graph method. Error images: $|I_i - f|$, for $i = 1, 2$, on a black background outside the region covered by the projection of the surface on the images. Errors have been magnified by a factor of 4 to make them visible.

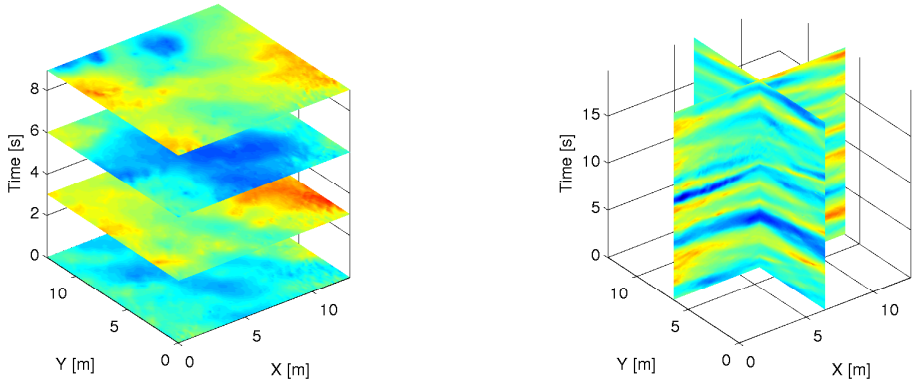


Figure 32: Left: Horizontal slicing (sequence of snapshots) of the wave space-time volume $Z(x, y, t)$. Right: Vertical slicing of the wave space-time volume $Z(x, y, t)$: reconstructions for fixed x-t or y-t planes. The wave heights Z have been pseudo-colored from blue (low) to red (high).

short surface waves. Long waves (low frequency components) may be attenuated by least-squares fitting of a multivariable, second-order polynomial surface to the wave heights. Low-pass filtering may also be necessary to remove measurement error in the high wave number region that is unphysical. We show results from the raw reconstruction and do not perform filtering unless otherwise stated. Windowing the height function prior to computing the 2-D spectrum reduces the effect of the discontinuities at the boundaries of the reconstructed region because the computation of the spectrum assumes periodicity of the input signal (Fourier analysis). Different window types have been tested in the literature and it has

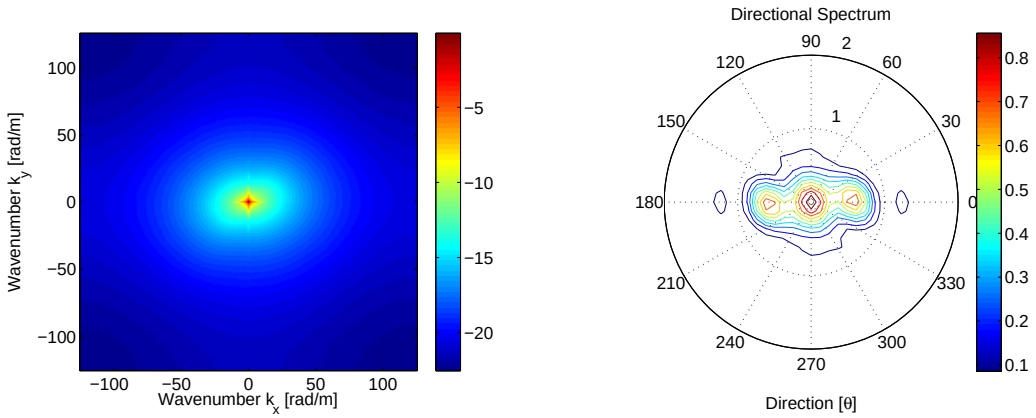


Figure 33: Mean directional spectrum (2-D Power Spectral Density) averaged over 2000 snapshots. Left: Spectrum on the frequency domain determined by the Nyquist frequency, in logarithmic scale. Right: Contour plot of the spectrum at low frequencies.

been shown that although different windows can affect the spectral energy level, they do not substantially alter the slope of the tail of the omni-directional wave number spectrum [90]. A commonly utilized window is the Tukey, or cosine-tapered, window. The mean directional (spatial) spectrum, averaged over all 2000 horizontal slices of the wave space-time volume V, is reported in Figure 33. The energy is clearly concentrated in low frequencies. The Nyquist wavenumber is $[k_x, k_y]_{\max} = [\pi/h, \pi/h] \approx [125.6\,\mathrm{rad}/m, 125.6\,\mathrm{rad}/m]$. The spectral resolution is given by $\Delta k_x = \Delta k_y = 2\pi/(Nh) \approx 0.123\,\mathrm{rad}/m$ for an $N = 2048$-point Fast Fourier transform (FFT) in each dimension.

The omni-directional spectrum is obtained from the directional spectrum of the ocean surface by a change to polar coordinates and integration over all possible directions. More specifically, if $\Phi(k_x, k_y)$ denotes the directional wave number spectrum (also known as the 2-D Power Spectral Density (PSD)) of the height function in a Cartesian coordinate system, where k_x (resp. k_y) is the wavenumber $k_x = \frac{2\pi}{\lambda_x}$ (rad/m) along the X axis (resp. $k_y = \frac{2\pi}{\lambda_y}$ along the Y axis), and $\Psi(k, \theta)$ is the same wave number spectrum in a polar coordinate system, the omni-directional spectrum $S(k)$, being k the wavenumber $k = (k_x^2 + k_y^2)^{1/2}$ and $\theta = \arctan(k_y/k_x)$ the wave direction, is defined as:

$$S(k) = \int_0^{2\pi} \Psi(k, \theta) k \, \mathrm{d}\theta.$$

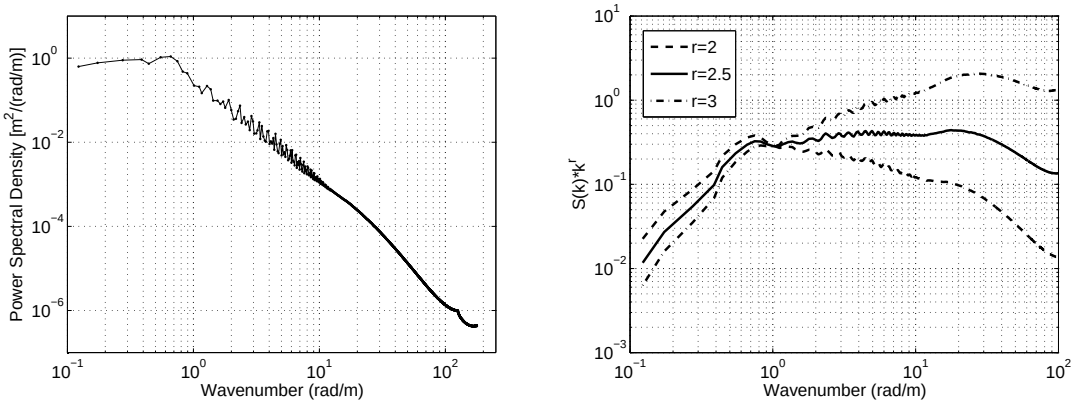


Figure 34: Left: Mean omni-directional spectrum $S(k)$ averaged over 2000 snapshots. Right: mean omni-directional spectrum multiplied by frequency variable: $S(k)k^r$ for $r = \{2, 2.5, 3\}$.

This definition differs from earlier ones [90] that do not include the multiplicative wavenumber (k) factor in the integrand. Due to symmetry of the directional spectrum caused by the fact that the elevation map is a real signal, the integration to compute $S(k)$ can be reduced to any contiguous interval of length π, e.g.,

$$S(k) = 2 \int_0^{\pi} \Psi(k, \theta) k \, \mathrm{d}\theta.$$

The mean omni-directional spectrum $S(k)$ corresponding to the directional spectrum of Figure 33 is reported in Figure 34. The tail of the omni-directional spectrum decays first as $k^{-2.5}$ and then the slope increases (in absolute value), in agreement with turbulence theory [112]. The plot on the right of Figure 34 clearly shows such behavior of the tail of the spectrum.

6.2.4.1 Sensitivity with respect to the weight of the shape regularizer

To show the sensitivity of the reconstructed surface with respect to the weight of the shape regularizer, two other reconstructions were carried with the variational graph method differing only in the value of α. The results are reported in Figure 35 in terms of the omni-directional spectrum. This experiment was performed to quantify the discussion in section 3.3 about the meaning of α as a controlling parameter for the amount of smoothness of the reconstructed surface shape. Observe the progressive behavior of the three reconstructions in Figure 35 at the high frequency region. The reconstruction with biggest α (0.2) is the

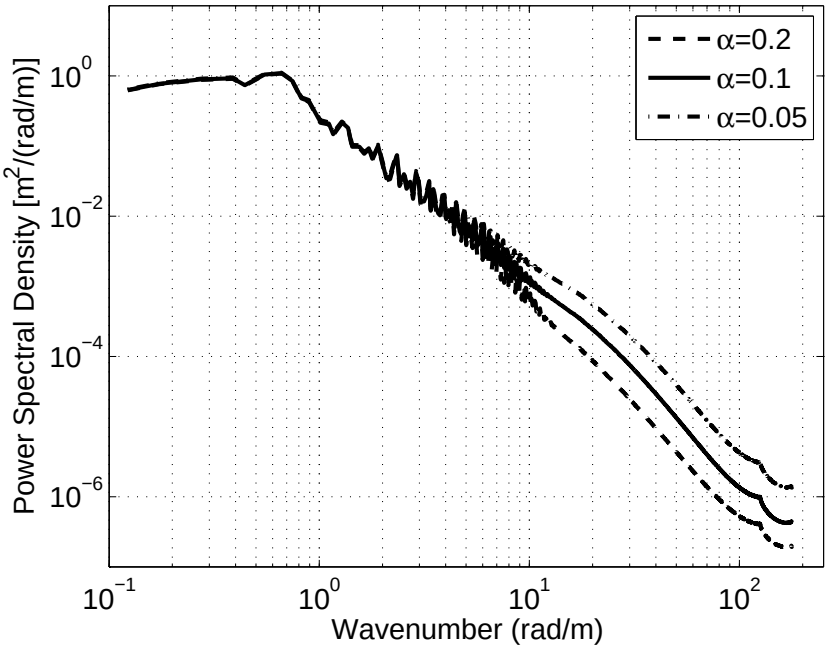


Figure 35: Mean omni-directional spectrum $S(k)$ averaged over 2000 snapshots for three different reconstructions according to the weight of the shape regularizer.

smoothest surface and this is represented in the steepest decay rate of the omni-directional spectrum. As α decreases, high frequency components rise, and therefore, the slope of the tail becomes milder.

6.2.4.2 Three-dimensional spectrum

More generally, Fourier analysis can be applied directly to the wave space-time volume $Z(x, y, t)$, resulting in a 3-D spectrum $\mathcal{Z}(k_x, k_y, \omega)$, where frequencies k_x, k_y and ω are the transformed variables of x, y and t, respectively. Recall that $\omega = 2\pi f$ is the angular frequency and $f \equiv f_t$ is the linear frequency, measured in Hertzs. The 3-D spectrum contains information of the propagation characteristics of the waves, such as their wavelengths, frequencies, and their directions and speeds of propagation.

From a practical point of view, the 3-D spectrum of the reconstructed $513 \times 513 \times 2000$ wave height grid is computed by averaging the 3-D spectrum of non-overlapping pieces of the grid. In our case, we decided to split the wave space-time volume along the temporal dimension to compute the 3-D spectrum on a Fourier grid with $512 \times 512 \times 512$ points; thus, each

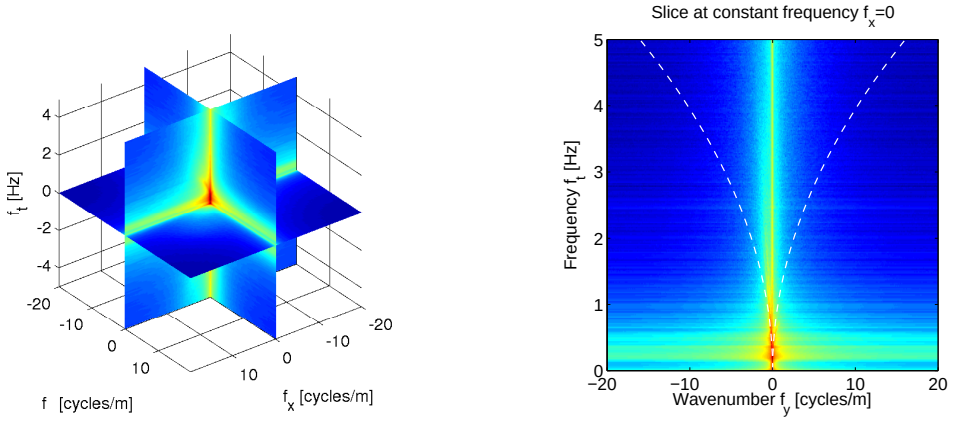


Figure 36: Left: Three orthogonal slices of the wave 3-D spectrum $\mathcal{Z}(k_x, k_y, \omega)$ through the frequency origin. Slices correspond to the values of $\mathcal{Z}(k_x, k_y, \omega)$ at planes $k_x = 0, k_y = 0$ and $\omega = 0$. Right: Vertical slice at $k_x = 0$ rad/m. Only half of the slice is displayed due to symmetry.

piece consists of $N_t = 512$ snapshots. The Nyquist wavenumbers are, in linear frequencies, $[k_x, k_y, f]_{\max} = [h^{-1}, h^{-1}, \Delta t^{-1}] = [20\,\text{cycles}/m, 20\,\text{cycles}/m, 5\text{Hz}]$. The spectral resolutions are given by $\Delta k_x = \Delta k_y = 1/(Nh) \approx 0.078\,\text{cycles}/m$ and $\Delta f = 1/(N\Delta t) \approx 0.02\,\text{Hz}$ for the 3-D FFT with $N = 512$-points in each dimension. Figures 36, 37 and 38 show the 3-D wave spectrum as well as its slices through the frequency axes. The white curve in the vertical slices corresponds to the linear dispersion relation in deep water:

$$k = \frac{\omega^2}{g}, \tag{95}$$

where g is gravity acceleration. Other researchers [15, 14] have measured the frequency-wave number spectrum $\omega - k$ for long wave ranges at nearshore events to estimate the surface currents and the water depth below the waves. Their measurements are also shown in comparison to the linear dispersion relation (95).

At the Crimean platform, the water depth is approximately 30 meters. Therefore, for all practical purposes with respect to our wavenumber resolution, the depth can be regarded as being infinite. A surface current vector has been estimated so that a "best" fit is obtained between the theoretical dispersion relation for linear gravity waves and the observed spectrum, as shown in Figure 39. The impact of currents on the spectrum can be found in [39, 53]. The current velocity vector is $\mathbf{u} \approx (-0.17, -0.45)$ m/s, with the dominant component in the y direction. This propagation direction agrees with the one observed

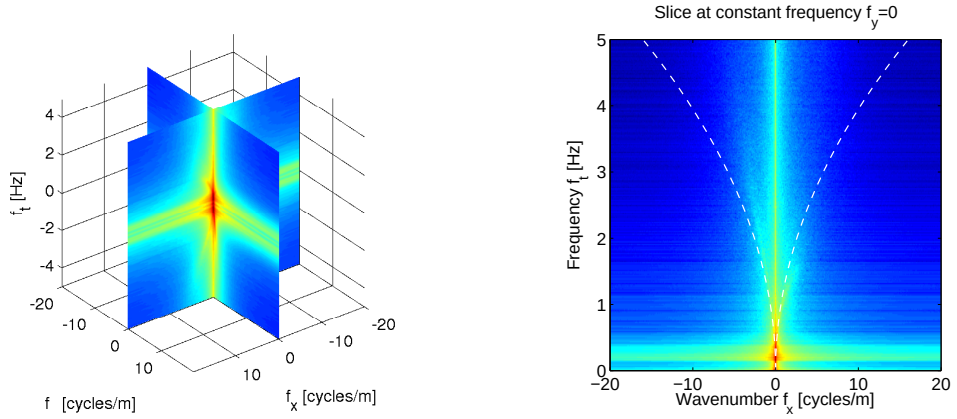


Figure 37: Left: Two orthogonal, vertical slices of the wave 3-D spectrum $\mathcal{Z}(k_x, k_y, \omega)$ containing the temporal frequency axis. Slices correspond to the values of $\mathcal{Z}(k_x, k_y, \omega)$ at planes $k_x = 0$ and $k_y = 0$. Right: Vertical slice at $k_y = 0$ rad/m. Only half of the slice is displayed due to symmetry.

by visual inspection of the stereo video data. Figure 39 shows strong physical evidence to support the hypothesis that the variational graph method presented in this thesis is capturing real waves propagating in the observed direction.

6.2.5 Time series at virtual probes. Statistical analysis

A sea state can be described by Fourier analysis as the linear superposition of a large number of elementary waves with distinct frequencies, amplitudes and random phases. Under constant wind conditions for short periods of time and regions of space, the resulting random process can be considered to be stationary in time and mean- and variance-ergodic in space. The rich content of the space-time reconstruction of the surface wave allows for the extraction of *time series* of wave displacements $Z_i(t) = Z(x_i, y_i, t)$ from the space-time volume V at virtual probes (x_i, y_i) in space, as illustrated in Figure 40. As is customary, we may regard a time series of wave heights $Z_0(t)$ as a realization of a random process. To characterize this random process, statistics are built upon elementary concepts of the waves within it. A wave is defined between two consecutive zero upcrossings. Within a wave we identify its crest (largest positive amplitude), its trough (largest negative amplitude), its wave length, etc. Several statistical and spectral parameters that characterize the sea states can be computed from such time series. Figure 41 gives an example of the extracted time series of wave heights, which contain a total number of roughly 1000 waves

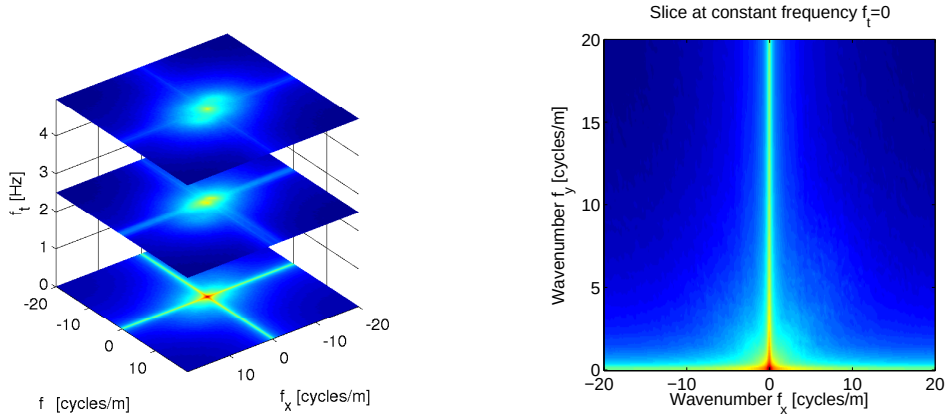


Figure 38: Left: Horizontal slicing of the wave 3-D spectrum $\mathcal{Z}(k_x, k_y, \omega)$. Each horizontal slice is the directional spectrum at different temporal frequency $\omega = 2\pi f$. Only half of the volume is displayed due to symmetry. Right: Horizontal slice (directional spectrum) at $f = 0$ Hz. Only half of the slice is displayed due to symmetry.

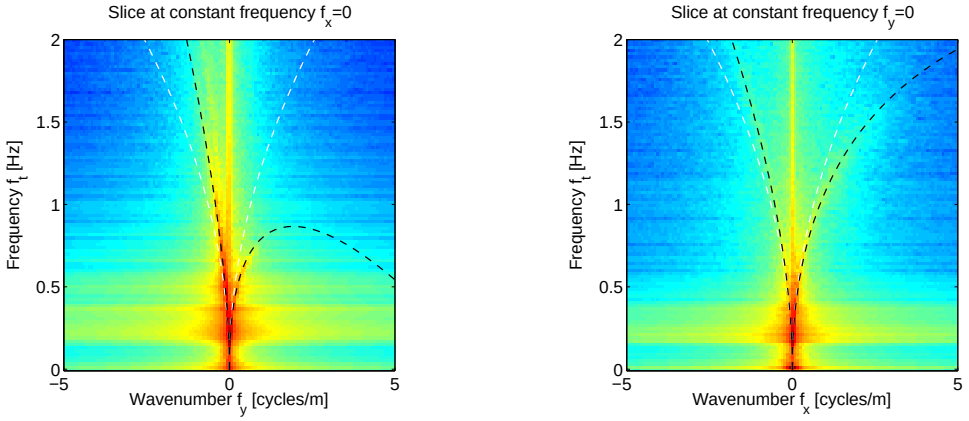


Figure 39: Vertical slices of the 3-D wave spectrum at frequencies $k_x = 0$ (left) and $k_y = 0$ (right). Superimposed on both plots: linear dispersion relation (white curve) and effect of a steady surface current with velocity vector $\mathbf{u} \approx (-0.17, -0.45)$ m/s (black curve).

The *significant wave height* is the parameter most commonly used to measure the strength of the wave motion, and it is defined by [5]

$$H_s = 4\sigma, \tag{96}$$

where σ is the standard deviation of the (zero-mean) random process, $\sigma^2 = \langle Z_0^2(t) \rangle$, being $\langle \cdot \rangle$ the temporal mean operator. The larger H_s, the higher the waves. Given the one-sided wave spectrum of the process, $S(\omega)$, the variance can be computed as the first moment of the spectrum, $\sigma^2 = m_0$, where $m_j = \int_0^\infty \omega^j S(\omega)\, \mathrm{d}\omega$ is the j-th spectral moment. Therefore, the significant wave height is also $H_s = 4\sqrt{m_0}$. For the time series in Figure 41, the significant

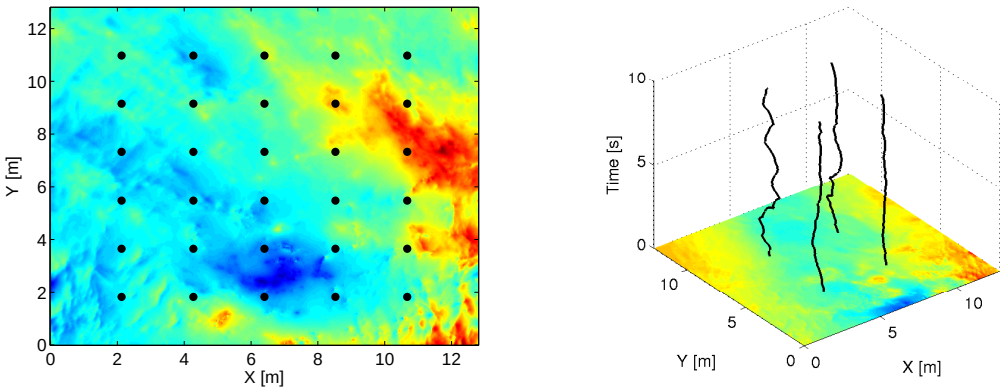


Figure 40: Left: Location of the virtual probes. Right: Illustration of extracted time series at probe points within the space-time volume $Z(x, y, t)$.

wave height is $H_s = 0.3$ m.

Rice's formula states that the expected number of h-upcrossings in a given large interval of length τ is

$$EX(h) = \frac{\tau}{2\pi}\sqrt{\frac{m_2}{m_0}}\exp\left(\frac{-h^2}{2m_0}\right), \tag{97}$$

which can be estimated from the wave spectrum. Since there is a one-to-one correspondence between zero upcrossings and waves, according to (97), the mean wave period is given by

$$\overline{T} = \frac{\tau}{EX(0)} = 2\pi\sqrt{\frac{m_0}{m_2}}. \tag{98}$$

The quantity $\overline{\omega} = \sqrt{\frac{m_2}{m_0}}$ can be regarded as the zero upcrossing angular frequency. For the time series in Figure 41, the mean wave period is $\overline{T} = 2.77$ s.

In the same figure, we show the observed Power Spectral Density estimated from time series extracted from the wave space-time volume. An FFT with 2048 points was used. Therefore, the spectral resolution is approximately $\Delta f = 5 \cdot 10^{-3}$ Hz. In a dispersive media like water, the physics of the waves states that, in deep water, the relationship between spatial and temporal frequencies is (95). Recall that $k = \frac{2\pi}{L}$ and $\omega = \frac{2\pi}{T}$ are the wavenumber and the angular frequency of a wave, respectively. The celerity or speed of a wave is

$$v = \frac{L}{T} = \frac{\omega}{k} \overset{(95)}{=} \frac{g}{\omega}.$$

Therefore, equation (95) means that longer waves (low frequency waves) travel faster than

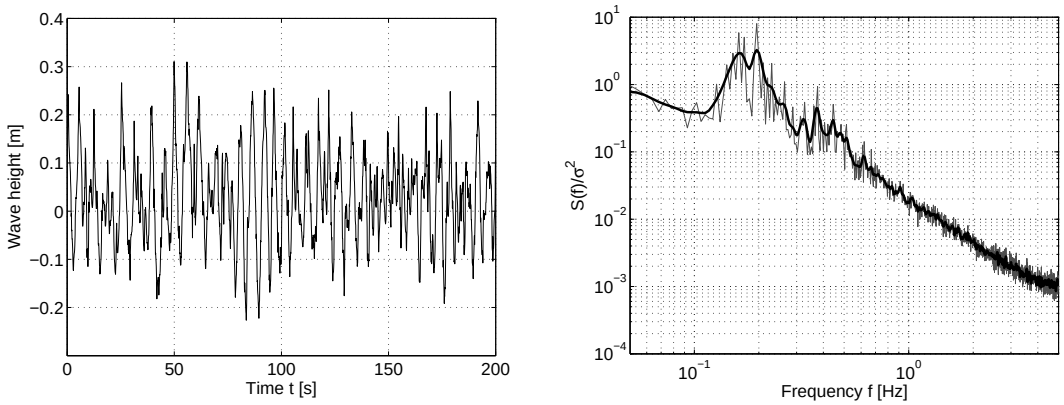


Figure 41: Left: Extracted time series of wave heights at a virtual probe point. Right: Frequency spectrum (Power Spectral Density, in units of m^2/Hz) averaged over all virtual probes. Note that the Nyquist frequency (half of the sampling frequency) is 5 Hz, according to the frame rate.

shorter waves (high frequency waves).

Given that the tail of the omni-directional wave number spectrum of the wave height decays according to $k^{-2.5}$ [112], one may use equation (95) to compute the corresponding decay rate in the temporal frequency spectrum. From $\tilde{F}(k)\,\mathrm{d}k = F(\omega)\,\mathrm{d}\omega$ and (95), if the tail of the wave number spectrum decays as $\tilde{F}(k) \propto k^{-2.5}$, the tail of the frequency spectrum decays as

$$F(\omega) \propto k^{-2.5}\frac{\mathrm{d}k}{\mathrm{d}\omega} = \left(\frac{\omega^2}{g}\right)^{-2.5}\frac{2\omega}{g} \propto \omega^{-4}.$$

Therefore, since the angular frequency ω and the ordinary frequency f are related by $\omega = 2\pi f$, the tail of the ordinary frequency spectrum decays as f^{-4}. This behavior is roughly observed in Figure 41.

Under the assumption of infinitely narrow spectrum, for large waves and large thresholds $h \gg \sigma$, there is a one-to-one correspondence between h-upcrossings and crests. Hence, the expected number of crests above h can be approximated by $EX(h)$, and the probability, within a wave, of having a crest larger than h (probability of exceedance) is given by

$$P(C > h) = \frac{EX(h)}{\overline{T}} = \exp\left(\frac{-h^2}{2m_0}\right). \tag{99}$$

Moreover, for Gaussian sea states, due to symmetry of crests and troughs, the wave height

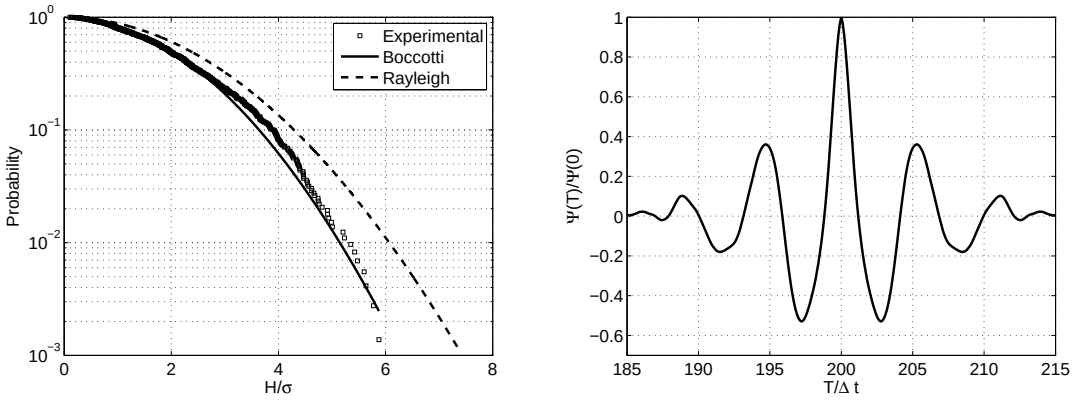


Figure 42: Left: Wave height exceedance probability estimated from all time series at virtual probes, compared to Rayleigh's distribution (100) and Boccotti's distribution (101). Right: Autocovariance function of a time series extracted from a virtual probe.

is twice the crest amplitude and (99) leads to Rayleigh's distribution:

$$P(\text{wave height} > H) = \exp\left(-\frac{H^2}{8m_0}\right), \tag{100}$$

which is one of the most common applied expressions to compute the probability of exceedance of the wave heights under the assumption of narrow spectrum. In real-life ocean waves, due to non-linearities, the probability of exceedance is not exactly Rayleigh distributed, but follows a modified distribution [98]. Both distributions will be plotted in Figure 42 after the autocovariance function is introduced.

Since the random process that models sea states is stationary, the autocovariance of a time series only depends on the lag between two times: $\psi(T) = \langle Z_0(t)Z_0(t+T)\rangle$. The autocovariance plays an important role in the theory of quasi-deterministic mechanics of wave groups [5]. The basic statistical properties of a sea state depend only on the characteristics of the first two waves of the autocovariance (the waves including the origin $T = 0$). For Gaussian sea states, the autocovariance is the inverse Fourier (cosine) transform of the wave spectrum $S(\omega)$.

By collecting the time waves observed at all the virtual probes indicated in Figure 40, one can estimate the wave height distribution, which is shown in Figure 42. A fair agreement with the Boccotti asymptotic form given by [5, 21]

$$P(\text{wave height} > H) \approx c\exp\left(-\frac{H^2}{4\sigma^2(1+\psi^*)}\right), \tag{101}$$

is observed. Here, the parameters c and $\psi^* \equiv |\psi(T^*)|$ both depend upon the first minimum of the wave covariance (see Figure 42). T^* is the abscissa of the absolute minimum of $\psi(T)$, which is assumed to be also the first local minimum for $T > 0$. In particular the mean values of c and ψ^* over the time series ensemble are $c \approx 1$ and $\psi^* \approx 0.52$.

For large waves ($h \geq 2\sigma$) satisfying the narrow spectrum assumption, one can further show that the expected shape of large waves is proportional to the covariance function:

$$\mu_C(T) = \overline{Z(t_0+T)|Z(t_0)=h} = \frac{h}{m_0}\psi(T), \tag{102}$$

and the deviation from such shape is given by the function

$$\sigma_C(T) = \sqrt{m_0 - \frac{\psi^2(T)}{m_0}}. \tag{103}$$

6.3 Conclusion

Empirical validation of the variational graph method and its different modifications presented in this thesis has been proved on real data. The variational graph method provides a dense and coherent representation of the reconstructed surface that is competitive with other variational disparity methods existing in the literature. Since there is no ground truth to establish a benchmark for comparison, a compelling validation has been given not only in terms of image-based criteria of the modeled images in comparison to existing reconstruction methods, but also in terms of physical properties that the reconstructed surface is known to satisfy from an ocean engineering point of view. The results of the different statistical tests and harmonic analysis carried out on the reconstructed space-time wave volume provide strong evidence to support the validation of the novel reconstruction approach presented in this thesis.

CHAPTER VII

INCORPORATING TEMPORAL COHERENCE

The variational reconstruction method for still images introduced in chapter 3 was naturally extended to process stereo video in chapter 6. Besides the sequential, snapshot-by-snapshot processing of stereo images, the variational framework developed in this thesis also allows for more ways to incorporate temporal coherence on the reconstructed surface.

Since the variational methods designed in this thesis for dynamic surfaces are tailored to a particular application (ocean waves), the ultimate goal of the reconstruction system is to include the spatial and temporal physics of the waves in the reconstruction step, i.e., in the original energy functional. Because ocean waves are governed by the wave equation, it would be desirable to include such law in the energy functional. This is however, a challenging problem. Before considering this approach, a natural way to enforce some temporal coherence of the reconstructed surface, besides the purely sequential processing above mentioned, is to include a simple temporal regularizer in our energy mode and solve the resulting problem. This is the object of this chapter and it implies a *simultaneous* estimation of the unknowns (height, radiance and possibly AGC coefficients) for *all* snapshots in a sequence. We call the resulting reconstruction a *manifold* reconstruction because the developed method estimates a manifold of graphs, also called wave heights or elevation maps.

7.1 Theoretical model

Consider the surface shape (height) and the radiance as functions of space and time (real time, not the fictitious time of gradient descent equations), i.e., as 3-D functions: $Z(u, v, \tau)$ and $f(u, v, \tau)$, with $(u, v) \in U$ and $t \in [0, T]$. Let us denote the new domain of Z and f by U_T, following the notation in Evans' book [16]. Let the symbol ∇ denote the gradient of Z

with respect to all variables, not only the spatial ones (u, v), i.e.

$$\nabla Z = (Z_u, Z_v, Z_\tau)^\top,$$

and similarly for ∇f. With this notation, the three variables u, v and τ are treated uniformly, as if all three of them were spatial variables. Finally, let the temporal sequence of images be denoted by $I_i(x, y, \tau)$, where $I_i(x, y, \tau_0)$ denotes the image acquired by the i-th camera at time τ_0.

Consider the following energy functional defined for a candidate sequence of surface shapes and radiances:

$$E(Z, f) = E_{\text{data}}(Z, f) + \alpha E_{\text{geom}}(Z) + \beta E_{\text{rad}}(f),$$

with weights $\alpha, \beta \in \mathbb{R}^+$. Let the data fidelity term, which measures the photo-consistency throughout the image sequence for a candidate sequence of wave heights, be

$$E_{\text{data}} = \int_0^T \sum_{i=1}^{N_c} \int_{\Omega_i} \phi_i \, \mathrm{d}\Omega_i \mathrm{d}\tau = \sum_{i=1}^{N_c} E_i,$$

with

$$E_i \doteq \int_0^T \int_{\Omega_i} \phi_i \, \mathrm{d}\Omega_i \mathrm{d}\tau,$$

and let the spatio-temporal regularizers on the smoothness of the surface shape and radiance be:

$$E_{\text{geom}} = \int_{U_T} \tfrac{1}{2} \|\nabla Z\|^2 \, \mathrm{d}\mathbf{u}_T = \int_0^T \int_U \tfrac{1}{2} (Z_u^2 + Z_v^2 + Z_\tau^2) \, \mathrm{d}\mathbf{u} \, \mathrm{d}\tau, \tag{104}$$

$$E_{\text{rad}} = \int_{U_T} \tfrac{1}{2} \|\nabla f\|^2 \, \mathrm{d}\mathbf{u}_T = \int_0^T \int_U \tfrac{1}{2} (f_u^2 + f_v^2 + f_\tau^2) \, \mathrm{d}\mathbf{u} \, \mathrm{d}\tau. \tag{105}$$

The new terms with respect to the still-image case, Z_τ^2 and f_τ^2, are the source of temporal coherence in the model. The data fidelity term corresponding to the i-th camera can also be expressed in the new integration domain U_T:

$$E_i = \int_0^T \int_U \phi_i \mathtt{J}_i \, \mathrm{d}\mathbf{u} \, \mathrm{d}\tau, = \int_{U_T} \phi_i \mathtt{J}_i \, \mathrm{d}\mathbf{u}_T,$$

where the Jacobian $\mathtt{J}_i$ (the geometric conversion factor between integrals) has the same expression as in the still-image case despite the fact that now the Jacobian is also a function of time τ (since $\mathtt{J}_i$ depends on the surface shape and the latter is a function of time).

Therefore, the composite energy can be expressed in the common and fixed domain of integration U_T:

$$E = \int_{U_T} L \, \mathrm{d}\mathbf{u}_T,$$

with Lagrangian

$$L(Z, \nabla Z, f, \nabla f, u, v, \tau) \equiv L = L_{\text{data}} + \alpha L_{\text{geom}} + \beta L_{\text{rad}}.$$

Definitions (26)-(29) are still applicable, but now the integrands take into account the temporal variable τ.

To minimize the proposed energy, let us compute the necessary optimality condition, which follows from the first variation (directional derivative) of the energy with respect to the unknowns:

$$D_{(h,w)}E(Z,f) \equiv \frac{\delta E}{\delta(Z,f)}(h,w) = \frac{\mathrm{d}}{\mathrm{d}\epsilon}E(Z+\epsilon h, f+\epsilon w)\bigg|_{\epsilon=0}.$$

Setting to zero the directional derivative for all admissible perturbations (h, w) we arrive at a coupled system of PDEs:

$$\begin{aligned}
L_Z - (L_{Z_u})_u - (L_{Z_v})_v - (L_{Z_\tau})_\tau &= 0 \qquad \text{in } U_T, && (106)\\
L_{Z_u}\nu^u + L_{Z_v}\nu^v + L_{Z_\tau}\nu^\tau &= 0 \qquad \text{on } \partial U_T, && (107)\\
L_f - (L_{f_u})_u - (L_{f_v})_v - (L_{f_\tau})_\tau &= 0 \qquad \text{in } U_T, && (108)\\
L_{f_u}\nu^u + L_{f_v}\nu^v + L_{f_\tau}\nu^\tau &= 0 \qquad \text{on } \partial U_T. && (109)
\end{aligned}$$

After calculations, similar expressions to those of the still-image optimality conditions are derived. That is why this approach is considered a natural extension of the aforementioned still-image case. Now, however, variables Z, f, and consequently $\mathtt{J}_i$ and I_i depend on the extra temporal variable τ:

$$\begin{aligned}
g(Z,f) - \alpha\Delta Z &= 0 \qquad \text{in } U_T, && (110)\\
b(Z,f) + \alpha\frac{\partial Z}{\partial \boldsymbol{\nu}} &= 0 \qquad \text{on } \partial U_T, && (111)\\
-\sum_{i=1}^{N_c}(I_i - f)\mathtt{J}_i(Z) - \beta\Delta f &= 0 \qquad \text{in } U_T, && (112)\\
\beta\frac{\partial f}{\partial \boldsymbol{\nu}} &= 0 \qquad \text{on } \partial U_T, && (113)
\end{aligned}$$

where the non-linear terms due to the data fidelity energy have the same expression as those of the still-image energy since the data fidelity energy does not depend on the new derivatives Z_τ, f_τ:

$$\begin{aligned} g(Z,f) &= \nabla_{\mathbf{u}} f \cdot \sum_{i=1}^{N_c} |\mathtt{M}^i| \tilde{Z}_i^{-3} (I_i - f)(u - C_i^1, v - C_i^2), \\ b(Z,f) &= \sum_{i=1}^{N_c} \phi_i |\mathtt{M}^i| \tilde{Z}_i^{-3} \big((u - C_i^1)\nu^u + (v - C_i^2)\nu^v \big), \end{aligned} \tag{114}$$

The 3-D Laplacians $\Delta Z = Z_{uu} + Z_{vv} + Z_{\tau\tau}$ and, similarly, Δf arise from the regularizing terms (104) and (105), respectively. $\partial * / \partial \boldsymbol{\nu} = (\nabla *) \cdot \boldsymbol{\nu}$ is the usual notation for the directional derivative along $\boldsymbol{\nu} = (\nu^u, \nu^v, \nu^\tau)^\top$, the normal to the integration domain U_T in the parameter space. $\nabla_{\mathbf{u}} f$ is the gradient of f with respect to the original spatial variables only. To simplify the model, we replace (111) by homogeneous Neumann boundary conditions, as in section 3.3.

Having introduced the manifold reconstruction model, let us make it more flexible by including an extra parameter $\rho^2 \geq 0$ to control the amount of temporal regularization of the solution relative to the amount of spatial regularization. The effect of this parameter is the substitution of the 3-D Laplacian $\Delta Z = Z_{uu} + Z_{vv} + Z_{\tau\tau}$ by the weighted sum $Z_{uu} + Z_{vv} + \rho^2 Z_{\tau\tau}$, and similarly for Δf. There are two possible interpretations to this anisotropic diffusion operator. The simplest one is that the operator arises by replacing the integrand of (104) by $\frac{1}{2}(Z_u^2 + Z_v^2 + \rho^2 Z_\tau^2)$, thus using a weighted norm instead of the Euclidean norm (and similarly for (105)). The second interpretation is that the anisotropic diffusion operator arises by using the Euclidean norm (104) in a deformed space where variable τ is scaled by ρ with respect to variables (u, v) to yield the desired non-uniform scaling of the gradients and the Laplacians, $\nabla Z, \nabla f, \Delta Z$ and Δf.

7.2 Numerical solution

7.2.1 Nested optimization and discretization

As in the still-image case (section 3.4), a nested minimization approach is adopted to address the solution of the optimality conditions. The linear PDE in the radiance is faster to solve using classical stationary iteration methods such as Jacobi or Gauss-Seidel rather than

setting up a gradient descent equation and using time-stepping[1] methods to solve it. The PDE in the height is more complicated, and a time-stepping method is used to solve it.

Numerical discretization of the PDEs is carried out using FDMs. Therefore, the integration domain U_T is discretized by means of a 3-D grid. Central differences are utilized for the Laplacians (both in spatial and temporal variables). Observe that the usual Laplacian stencil is no longer 5-point, but 7-point because each 3-D grid point has 6 adjacent neighbors. That is, the 3-D version of formula (47) is, taking into account the controlling parameter ρ^2,

$$\begin{aligned}(\Delta f)_{i,j,k} \approx \frac{1}{h^2}\big(&v_{i+1,j,k} + v_{i-1,j,k} + v_{i,j+1,k} + v_{i,j-1,k} - 4v_{i,j,k} \\ &+ \tilde{\rho}^2\left(v_{i,j,k+1} + v_{i,j,k-1} - 2v_{i,j,k}\right)\big),\end{aligned} \tag{115}$$

where $\tilde{\rho} = \rho\, h/\Delta\tau$ and $v_{i,j,k}$ is an approximation to the exact solution $f(x_i, y_j, \tau_k)$, with $\tau_k = k\Delta\tau$. The quantity $\tilde{\rho}$ can be interpreted as the step ratio that states the relationship between the grid steps $h = \Delta u = \Delta v$ and $\Delta\tau$ in the anisotropic space previously mentioned.

7.2.2 3-D multigrid solver

Since the three variables u, v, τ are uniformly treated as space variables, to solve the above optimality PDEs we consider an extension of the spatial, 2-D multigrid method used for the still-image problem. Thus, now each multigrid level consists of a 3-D grid instead of a 2-D grid. Inter-grid transfer operators such as prolongation and restriction are defined for these 3-D grids. Trilinear interpolation and 3-D full weighting are the corresponding extension from their 2-D counterparts for the prolongation and restriction operators, respectively. The description of the 3-D inter-grid transfer operators can be found in [103]. Each relaxation iteration implies the update of the values of the approximate solution (either Z of f) at *all* the grid points, that is, all physical time slices of the 3-D grid are updated in the same iteration. This simultaneous processing property is the main difference with respect to the sequential scheme.

Due to the extra dimension added to the problem, memory becomes a precious resource. Storage is required not only for the unknowns Z and f at the different levels of multigrid,

[1] artificial time, not the physical time τ.

but also for the coefficients of the linear/linearized PDEs and other temporary variables. In-place updates of the unknowns are preferred over the alternative (e.g., Gauss-Seidel vs Jacobi iteration) because the former allow longer sequences of images to be processed with the same amount of memory.

Full Multigrid with zero initial condition is used to initialize the 3-D multigrid solver. It is also possible to initialize it with the solution from a fast sequential 2-D multigrid, possibly computed using a lower resolution version of the problem.

7.2.2.1 Comparison to sequential 2-D multigrid

To gain intuition on the sequential and manifold reconstruction algorithms introduced in this thesis, we briefly discuss some of their differences from the point of view of their multigrid solvers. Both share a common theoretical modeling, but the major difference between them is that sequential reconstruction relies on 2-D multigrid (snapshot-wise), whereas manifold reconstruction is based on 3-D multigrid. Although speed is not the focus during the development stage of the newly investigated manifold reconstruction method, we discuss this topic of practical relevance by point out reasons in favor and against manifold reconstruction being faster than sequential reconstruction.

Comparison of inter-grid transfer operators in 2-D and 3-D. On the one hand, 3-D restriction is faster than sequential 2-D restriction because each coarser level of 3-D multigrid has roughly half the number of snapshots of the previous level, whereas in sequential 2-D multigrid the number of snapshots remains constant and maximal for all coarse levels. On the other hand, assuming full weighting restriction, each point in the coarse grid requires a weighted sum of 9 points in case of 2-D restriction and 27 points in case of 3-D restriction. Thus, each point-wise 3-D restriction involves more neighbors and calculations than each 2-D restriction. Similar comments apply to the prolongation/interpolation operators: 3-D prolongation using trilinear interpolation is computed in a smaller number of grid points than sequential 2-D prolongation by bilinear interpolation, but each point-wise operation of trilinear interpolation is more expensive than each bilinear interpolation operation. For details on 2-D and 3-D inter-grid transfer operators, the reader is encouraged to review [103].

Sequential vs. manifold processing. The update of each grid point in manifold reconstruction is slower than the update of each point in sequential reconstruction because the former depends on more neighbors: it requires the calculation of temporal derivatives that are not present in sequential processing, and the 3-D Laplacian formula is also more involved than the 2-D Laplacian. In addition, sequential processing only performs full multigrid for the reconstruction of the first snapshot (and one-sided propagation of the solution for the initialization of the remaining snapshots), whereas manifold processing carries out full multigrid for all snapshots. Moreover, the artificial time step to evolve the non-linear gradient descent PDE in the height (e.g. formula (54)) is smaller in manifold processing than in sequential processing because the former takes the minimum time over all snapshots, whereas the latter takes a different the minimum over each snapshot. This implies that manifold processing requires more iterations than sequential processing to reach the same final, artificial time. On the other hand, manifold processing is more robust than sequential processing because it is less prone to be trapped in a local minimum caused by the poor reconstruction of a snapshot. Errors are distributed better among all frames in manifold processing than in sequential processing, whereas the propagation of errors can lead to totally wrong reconstructions using sequential processing. In general, we may conclude in a first analysis that sequential reconstruction is simpler and faster but more sensitive to noise than manifold reconstruction.

7.3 Extended radiance model

If AGC correction (section 3.6) is also included in the radiance model, it is sensible to assume that the gain of the cameras does not change during the (short) sequence of images. Therefore, only 2 coefficients per camera are used (for all the images in the sequence).

The radiance model with AGC compensation replaces f with the new radiance model

$$f \leftarrow a_i f + b_i, \tag{116}$$

where $i \in \{1, 2, \ldots, N_c\}$ is the camera index. Assuming a quadratic photometric criterion, the AGC correction (116) implies the new criterion $\phi_i = \frac{1}{2}(I_i - (a_i f + b_i))^2$. Setting $i = 1$ to be the reference camera ($a_1 = 1, b_1 = 0$), the necessary optimality conditions with respect

to the AGC parameters are, for $2 \leq i \leq N_c$,

$$\frac{\partial E}{\partial a_i} = 0 = \frac{\partial E}{\partial b_i}.$$

Since ϕ_i is quadratic in the intensities being compared, its derivative is linear and the necessary optimality conditions become $N_c - 1$ linear systems of equations of the form

$$\begin{pmatrix} \int_0^T \int_{\Omega_i} f^2 \, \mathrm{d}\hat{\mathbf{x}}_i \, \mathrm{d}\tau & \int_0^T \int_{\Omega_i} f \, \mathrm{d}\hat{\mathbf{x}}_i \, \mathrm{d}\tau \\ \int_0^T \int_{\Omega_i} f \, \mathrm{d}\hat{\mathbf{x}}_i \, \mathrm{d}\tau & \int_0^T \int_{\Omega_i} \mathrm{d}\hat{\mathbf{x}}_i \, \mathrm{d}\tau \end{pmatrix} \begin{pmatrix} a_i \\ b_i \end{pmatrix} = \begin{pmatrix} \int_0^T \int_{\Omega_i} I_i f \, \mathrm{d}\hat{\mathbf{x}}_i \, \mathrm{d}\tau \\ \int_0^T \int_{\Omega_i} I_i \, \mathrm{d}\hat{\mathbf{x}}_i \, \mathrm{d}\tau \end{pmatrix},$$

which are evaluated in the parameter space:

$$\begin{pmatrix} \int_{U_T} f^2 \mathtt{J}_i \, \mathrm{d}\mathbf{u}_T & \int_{U_T} f \mathtt{J}_i \, \mathrm{d}\mathbf{u}_T \\ \int_{U_T} f \mathtt{J}_i \, \mathrm{d}\mathbf{u}_T & \int_{U_T} \mathtt{J}_i \, \mathrm{d}\mathbf{u}_T \end{pmatrix} \begin{pmatrix} a_i \\ b_i \end{pmatrix} = \begin{pmatrix} \int_{U_T} I_i f \mathtt{J}_i \, \mathrm{d}\mathbf{u}_T \\ \int_{U_T} I_i \mathtt{J}_i \, \mathrm{d}\mathbf{u}_T \end{pmatrix},$$

and whose solution is straightforward. Compared to the still-image case, the new coefficients of the linear system have an extra integration across time, as expected since the coefficients are the same for all images in the sequence.

The radiance model with AGC compensation affects the EL equations in the same way as in the still-image case, according to the substitution (116), and similarly with the gradient, $\nabla_{\mathbf{u}} f \leftarrow a_i \nabla_{\mathbf{u}} f$.

In the multigrid framework that implements the solution of the manifold reconstruction problem, the AGC coefficients are computed at the finest level available and are used in all (coarser) levels and time indices τ to maintain consistency of the model in the multiresolution setup.

7.4 Experiments

Numerical experiments for the manifold reconstruction method have been carried out on the Crimea sequence of stereo images mentioned in previous chapters. Two cameras ($N_c = 2$) provide images of size 1624×1236 pixels, acquired at a sampling rate of 10 Hz.

7.4.1 Reconstruction size and resolution. Memory constraints

Current implementation of the manifold reconstruction method requires all relevant information (input images, processing grids, etc.) to be stored in the computer's main memory

(Random Access Memory - RAM), which is directly accessible to the Central Processing Unit (CPU). Thus, the limit in the amount of available main memory in the computer performing the reconstruction imposes a constraint on the size of the achievable manifold reconstruction. For example, for a 6-level multigrid solver and the spatial resolution used in the experiments of the sequential processing method of chapter 6, i.e., a spatial grid with 513×513 points ($h = 2.5$ cm), this limiting argument implies that the manifold reconstruction method of the current chapter can at most handle roughly[2] 65 snapshots in a computer with 2 GBytes of RAM.

For a fixed amount of memory, more snapshots can be handled by the manifold reconstruction method if the spatial resolution of the physical grid is decreased because there is freedom in the choice of the number of grid points placed in each of the dimensions of the 3-D grid to satisfy a total grid size constraint. Therefore, there is a trade-off between the spatial and temporal resolutions that the manifold method can handle for a fixed size of the computational grid: one can either reconstruct a few snapshots with high spatial resolution or a larger number of snapshots at a coarser spatial resolution.

To simultaneously reconstruct a decent amount of snapshots, we decrease the spatial resolution by 4, from $h = 2.5$ cm to $h = 10$ cm, which allows us to increase the number of snapshots being processed by a factor of $4^2 = 16$, up to 1025 snapshots. Smaller images at the appropriate scale are fed to the manifold reconstruction algorithm to be consistent with the details that the resolution of $h = 10$ cm allows the algorithm to capture. Thus, the input data is coarsified/down-sampled via scaling of the images by a factor of 4 in each dimension (images of size 406×409 pixels). The manifold reconstruction algorithm operates on the entire sequence of images by reconstructing pieces (sub-sequences) of 1025 consecutive snapshots. In each sub-sequence, the method yields a reconstruction of wave heights on a discretized grid of $129 \times 129 \times 1025$ points. In summary, the spatial and temporal resolutions are $h = 10$ cm and $\Delta\tau = 0.1$ s, respectively. No temporal snapshot decimation is carried out. Other coarse-to-fine strategies are also possible to ensure a smooth transition near the temporal boundaries of the sub-sequences.

[2]The number of snapshots is chosen in powers of 2 plus one to be suitable for the multigrid method.

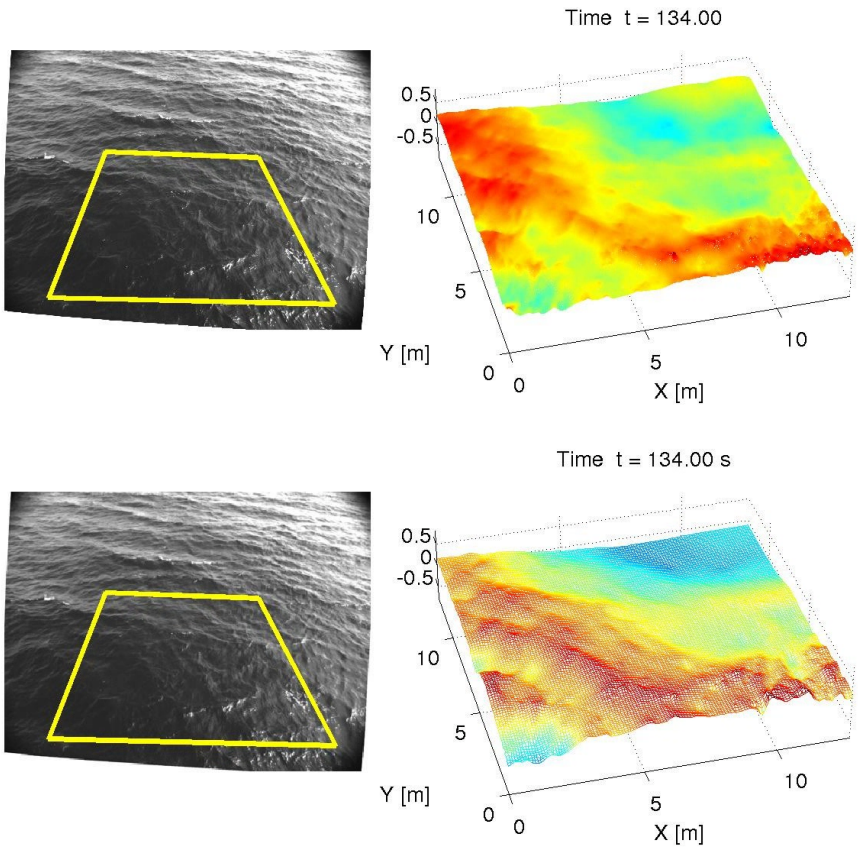


Figure 43: Right: reconstruction of a snapshot at spatial resolutions $h = 2.5$ cm (top) and $h = 10$ cm (bottom). Left: one of the two input images at the corresponding resolution (top: 1624×1236 pixels, bottom: 406×309 pixels), with highlighted reconstructed region of interest.

An example of the reconstructions obtained at the two spatial resolutions previously considered is given in Figures 43 and 44. These reconstructions have been generated by the sequential method. Nevertheless, they show that the reconstruction at the low resolution captures the gist (i.e., low spatial frequency components) of the wave heights, thus resembling the reconstruction at the high spatial resolution.

Snapshots can also be decimated in time. The linear dispersion relation, which relates spatial and temporal frequencies of waves in deep water (95), serves as a physical criterion to choose a reasonable frame rate for a given spatial resolution, and vice versa. For instance, in the Crimean example with grid resolution $h = 10$ cm, assuming the minimum spatial wavenumber that the algorithm reconstructs reliably is $\lambda = 4h$, the corresponding frequency

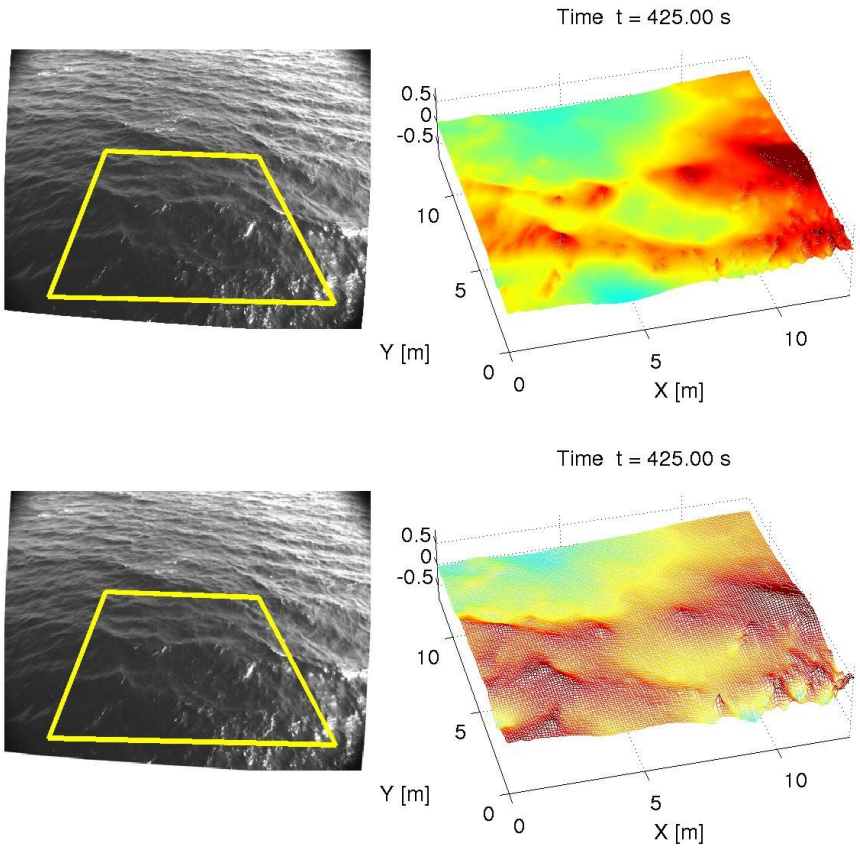


Figure 44: Right: reconstruction of another snapshot at spatial resolutions $h = 2.5$ cm (top) and $h = 10$ cm (bottom). Left: one of the two input images at the corresponding resolution (top: 1624×1236 pixels, bottom: 406×309 pixels), with highlighted reconstructed region of interest.

of the wave is

$$f = \frac{\omega}{2\pi} \overset{(95)}{=} \frac{1}{2\pi}\sqrt{g\frac{2\pi}{\lambda}} = \sqrt{\frac{g}{2\pi\, 4h}} \approx 2 \text{ Hz}.$$

If this is the maximum temporal frequency that the algorithm reconstructs reliably, the corresponding Nyquist rate (minimum sampling rate required to avoid aliasing) is twice as much, 4 Hz, meaning that snapshots should be at most $\Delta\tau \leq 1/4 = 0.25$ seconds apart. In the experiments, we used a smaller $\Delta\tau = 0.1$ s since the acquisition rate (10 Hz) allowed so, but we could have used $\Delta\tau = 0.2$ s (by picking one every two snapshots, i.e., temporal snapshot decimation by a factor of 2) to achieve results with similar validity and interpretation, but expanding to twice the physical time interval. Thus, the linear dispersion relation can be used at first steps of the reconstruction process to decide rough estimates of the spatial and temporal resolutions for the manifold reconstruction method.

Figure 45: A slice at constant $u = u_0$. Top: surface height $Z(u_0, v, \tau)$ (grayscale encoded). Bottom: surface radiance $f(u_0, v, \tau)$. Horizontal axis is time τ.

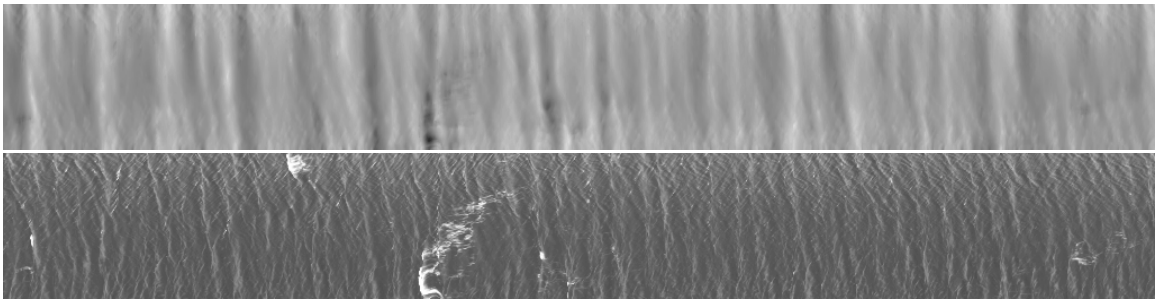


Figure 46: A slice at constant $v = v_0$. Top: surface height $Z(u, v_0, \tau)$ (grayscale encoded). Bottom: surface radiance $f(u, v_0, \tau)$. Horizontal axis is time τ.

7.4.2 Comparison to the sequential reconstruction method

Experiments were carried out using manifold reconstruction for a sequence of 4100 consecutive snapshots, split in sub-sequences of 1025 snapshots. A 6-level full multigrid method with 1000 iterations per level and 1500 iterations at the finest level was performed, with $NV_Z = NV_f = 2$ V-cycles per iteration (Figure 9) and one pre- and post-relaxation sweeps per level. The initial manifold surface consisted of the zero height surface (for all snapshots). Different values of the parameter ρ were tested: $\tilde{\rho} = \rho h/\Delta\tau = \{0.1, 0.2, 0.5, 1\}$. Observe that the temporal coherence of the manifold reconstruction decreases as $\rho \to 0$. In the limit, $\rho = 0$ is equivalent to the reconstruction of each snapshot independently, using full multigrid on each of them, but with a common artificial time step shared among all snapshots.

Figures 45-47 show slices of the surface height and radiance functions obtained by the manifold reconstruction method. The computational grid has $129 \times 129 \times 1025$ points, with spatial and temporal resolutions $h = 10$ and $\Delta\tau = 0.1$ s, respectively. In particular,

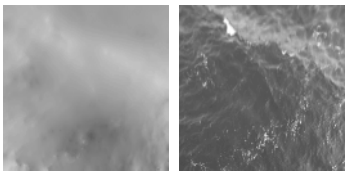


Figure 47: A slice at constant $\tau = \tau_0$. Size: 129×129 grid points. Left: surface height $Z(u, v, \tau_0)$ (grayscale encoded). Right: surface radiance $f(u, v, \tau_0)$. Images are displayed small on purpose to match the size of the slices in Figures 45 and 46.

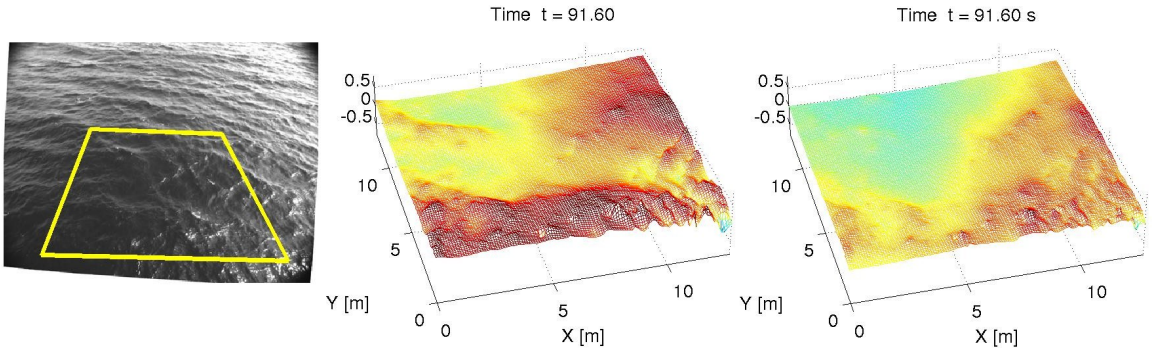


Figure 48: Manifold reconstruction at one of the snapshots. Left: one of the coarse input images of the stereo pair; Center: reconstructed wave height with $\tilde{\rho} = 0.1$; Right: reconstructed wave height with $\tilde{\rho} = 1.0$.

the results correspond to the case $\tilde{\rho} = 0.1$. Observe the oscillating patterns of the ocean waves in both the height and radiance functions of the u and v slices (Figures 45 and 46). The radiance function also captures the location and motion of sea foam caused by breaking waves. Figure 47 gives the reconstructed surface height and radiance for a particular snapshot of the sequence, as already seen in, for example, Figures 11 and 24.

By visual inspection of the reconstruction (see, for instance, Figure 48), one concludes that the values $\tilde{\rho} = \{0.5, 1\}$ are too large: temporal derivatives are penalized too much with respect to spatial derivatives, yielding a reconstructed surface shape that is very smooth in time and does not capture the wave patterns present in the stereo video data. Drawing an analogy with linear signal processing, the anisotropic diffusion carried out by the (weighted) Laplacian operator has a low-pass filtering effect: it limits the temporal bandwidth of the output signal, thus reducing noise but also destroying the desired wave signal.

Table 5 compares the photometric error of the reconstructed sequence of wave heights obtained by means of the sequential method and the manifold method with different values

Table 5: Comparison of photometric error (92) for several methods used to reconstruct the same coarse stereo image sequence: the variational graph sequential method (chapters 3 and 6) and the variational graph manifold method (chapter 7).

Method	mean	standard deviation
Sequential	3.613	0.398
Manifold, $\tilde{\rho} = 0.1$	3.615	0.393
Manifold, $\tilde{\rho} = 0.2$	3.621	0.391
Manifold, $\tilde{\rho} = 0.5$	3.657	0.392

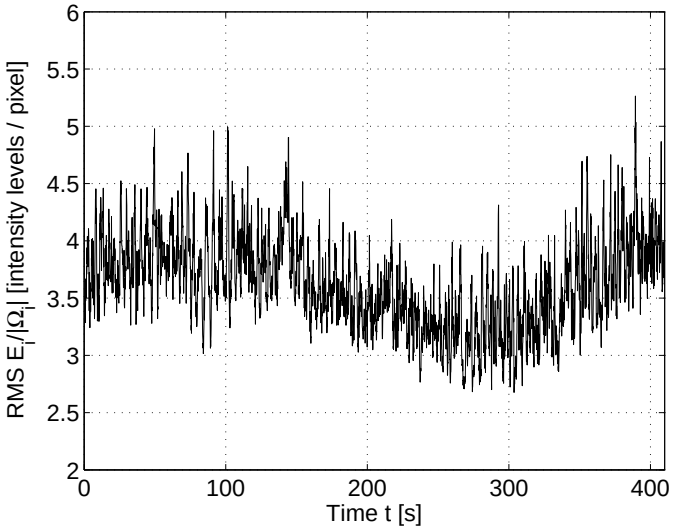


Figure 49: Photometric error for manifold reconstruction ($\tilde{\rho} = 0.1$) at coarse dataset.

of the parameter ρ. Observe that the magnitude of the photometric error is small and very similar for all values of ρ, however the differences in reconstructed surface shape (wave heights) are more noticeable by means of other methods such as the statistics of the wave heights. The particular profile of one of the methods is shown in Figure 49. The other methods follow very similar profiles, as it can be inferred from the small difference in their photometric statistics (Table 5).

The spectral and statistical analysis carried out in chapter 6 can be repeated for the manifold method to validate the resulting reconstruction. However, no significant differences with respect to the sequential reconstruction method are reported except for the fact that

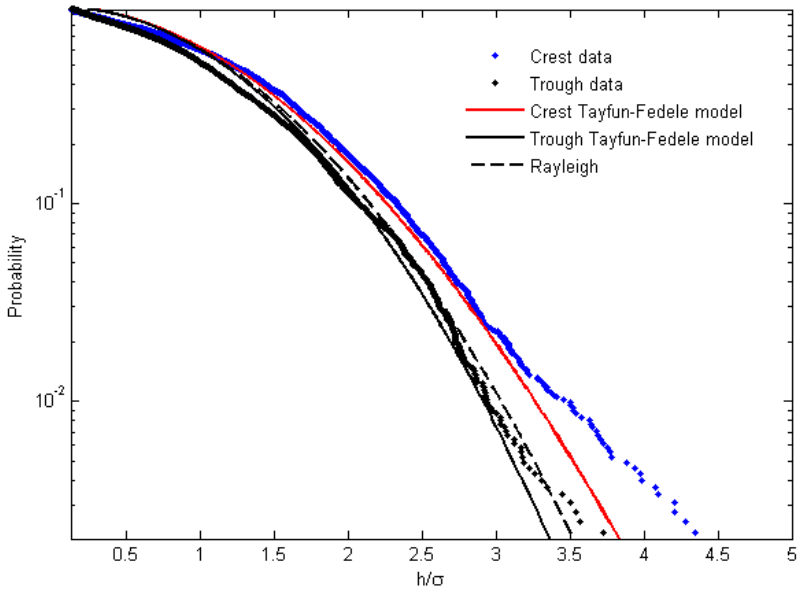


Figure 50: Empirical exceedance probabilities of wave crests and troughs from time series extracted at virtual probes of the manifold reconstruction with $\tilde{\rho} = 0.1$. Rayleigh and Tayfun-Fedele theoretical models are also shown for comparison.

the manifold reconstruction method shows an improvement in the estimation of the crest-trough asymmetry of wave heights (Figure 50).

7.5 Conclusion

In this chapter, the variational graph manifold method for the space-time coherent reconstruction of ocean waves has been discussed and developed. Due to the convenient representation of the ocean surface as a height function, the incorporation of a temporal dimension is straightforward. This benefit of the graph representation makes the reconstruction problem significantly simpler than what it would be if the level set framework was used instead. The manifold method has been tested on stereo video data from real ocean waves at an off-shore platform in the Black Sea. The qualitative and photometric performance of the method has been demonstrated on a coarser version of the dataset due to memory limitations of the implementation with respect to the size of the original dataset. By design, the manifold reconstruction method is more robust than the sequential method, at the expense of speed. In addition, it captures better some physical properties of wave heights than the sequential method. This evidence shows that the incorporation of the physics and coherence

in our variational method produces tangible improvements and encourages us to continue with this line of research to achieve better results. These insights justify the research on the incorporation of the wave equation in the reconstruction process. Finally, the manifold view developed in this chapter can also be applied to the variational disparity method in [3]. This topic may be further investigated in the future.

CHAPTER VIII

CONCLUSIONS AND FUTURE WORK

We have designed and implemented a novel variational stereo method and framework for the spatio-temporal reconstruction of dynamic surfaces such as ocean waves. The primary purpose in considering the design of this variational method was to obtain accurate measurements of ocean waves to enrich the understanding of sea states. However, the method has been formulated in a framework as generic as possible to extend its applicability to other surfaces besides ocean waves. The novel remote-sensing observational technology based on our variational method has a broad impact in the design of off-shore structures.

We have, in chapter 3, developed an object-based variational stereo method for the three-dimensional reconstruction of surfaces from multi-view still images. We have developed a simple but complete generative model of the images that takes into account both the surface of interest in the scene and the camera parameters. This generative model is object-centered. More specifically, it is adapted to surfaces that admit a representation in the form of a graph and that "support" a smooth radiance function. This model has been designed to overcome the theoretical and computational shortcomings of previous stereo methods discussed in chapter 2, both classical and variational, for the reconstruction of smooth surfaces such as ocean waves. We have cast the reconstruction problem into a variational optimization process in which the unknown surface shape and radiance of the interest object in the scene are obtained as the minimizer of an energy functional. The energy functional combines ideas from the Stereoscopic Segmentation method of Yezzi and Soatto [110], and variational 3-D reconstruction of Faugeras and Keriven [19]. We have shown from several points of view, the connection between the unconstrained surface evolution dictated by the gradient of the energy functional for compact surfaces and the constrained surface evolution for surfaces in the form of a graph that are used in our physical model of ocean waves.

We have rigorously derived the associated optimality (Euler-Lagrange (EL)) equations

that govern the model and explained the implications of the different terms that arise in the equations: the lack of image derivatives that provides improved robustness to image noise, the incorporation of the necessary visibility condition in the Jacobian of the transformation between the image domain and the parameter space of the surface, the implications of the choice of photometric criterion and its derivatives on the EL equations and stability condition of the numerical solvers, etc. Numerically, we have analyzed the complexity of different solvers and opted for the most flexible and efficient ones, multigrid methods, to yield an approximation to the solution in a reasonable amount of time. We have developed, using multigrid methods, a nested optimization approach to solve the coupled system of non-linear EL equations.

In chapter 4, we have developed, under the assumption of a calmed sea climate, a solution to the crucial problem of estimating the gravity plane for ocean waves. We have shown how the same continuous modeling principles used in variational problems, paired with multi-view geometry, e.g., planar homographies, can be applied to design an integral cost function whose minimizer is the gravity plane.

Utilizing variational principles, we have shown, in chapter 5, that it is possible to include global, physical properties of ocean waves in the method explained in chapter 3. These properties have been added in the form of weak constraints on the unknowns by means of energy penalties. We have demonstrated that the area-based regularizer naturally pairs with a geometric inner product and that this new shape regularizer only affects the diffusive component of the non-linear PDE in the height function. Statistical regularizers have also been developed to weakly enforce global, empirically-based wave height distribution models in the reconstruction step. Two different approaches (Cumulative Distribution Function (CDF) and Probability Density Function (PDF)) have been derived. These approaches appear as new non-linearities in the coupled system of PDEs. Also, we have demonstrated that a sub-quadratic photometric criterion improves the robustness of the variational method to isolated specular regions and abrupt intensity changes in the surface radiance caused by breaking waves and/or sea foam. This same effect can also be accomplished by a logarithmic pre-processing of the stereo images.

In chapter 6 we have shown how the variational graph stereo method developed in chapter 3 for still images can be easily extended to account for temporal processing of image sequences. Experimental evidence has been given to support the claim that our variational method yields competitive results with respect to other existing reconstruction methods. Lacking of ground truth, the results of the developed method have also been validated according to empirically-based physical models currently used in ocean engineering standards. Finally, in chapter 7, we have developed the most natural extension of the variational stereo method for still images to enforce temporal coherence during the reconstruction of image sequences. This lead us to the manifold reconstruction method.

From the previous remarks, we conclude that the main research objectives stated in section 1.3 have been satisfied. This thesis has opened up a new line of research in the application of variational stereo methods to the ocean engineering problem of acquiring measurements of ocean waves. We have established the mathematical foundation on how to combine different techniques (image processing, computer vision, PDEs, etc.) in a variational framework to address the aforementioned problem. The methodology developed in this thesis is suitable for the exploration of new research directions due to the flexibility and broad applicability of both the theoretical variational framework and the numerical multigrid solvers.

Encouraged by the significant impact that the empirically-based statistical regularizers developed in chapter 5 have on the shape of the reconstructed surface, we predict that the most natural suggested research direction consists of the extension of the theoretical ideas presented in this thesis to the exploration of ways to incorporate mechanical properties of the ocean surface in the variational formulation of the reconstruction problem. Ideally, the final goal would be to merge an empirically-justified non-linear version of the wave equation that governs ocean waves into the reconstruction step, that is, to estimate the space-time surface in the scene that best fits the observed video data subject to the fact that the surface satisfies the wave equation. This means that, using similar ideas to those in chapter 7, the surface reconstruction problem is cast as the problem of segmenting a three-dimensional manifold of graphs embedded in a four-dimensional space (space-time

continuum), without counting the radiance function. This is an ambitious goal since the resulting model becomes fairly complex both theoretically and numerically, and it is not yet clear what is the best way to merge such dynamical, physical properties. Two direct alternatives arise: (*i*) the weak enforcement of the dynamical constraint by means of extra penalty terms in the energy functional of the problem, and (*ii*) the strong enforcement of the constraint via Lagrangian multipliers and adjoint PDE operators. In the first case, a Hamiltonian formulation of the wave equation may be the most compelling approach from a physical point if view. A penalty on the Hamiltonian of the candidate surface leads to a global evolution law (the perturbation of a grid point in one iteration depends on more than its direct neighboring points) that tends toward becoming local as the candidate solution approaches the minimizer of the penalty (because the wave equation can be described as local perturbations of the surface).

Another interesting approach to examine toward the assimilation of the wave equation in the presented generative model is that of spectral methods [68, 101], where the solution to the EL equations of the arising energy functional is sought using efficient algorithms in the Fourier (frequency) domain. Nevertheless, there is a trade-off in this approach: the formulation of the wave equation is conveniently expressed in the frequency domain, but the formulation of already derived terms, such as the data fidelity term or the shape regularizer, is not. Due to the Uncertainty Principle in the localization of a function and its Fourier transform, a local evolution law in the spatial domain, such as that derived from the data fidelity term, becomes global in the frequency domain, and vice versa. Either if the surface evolution is carried out in the frequency domain, by transforming already derived gradient terms, or if the evolution is carried out in the spatial domain, by transforming the gradient term corresponding to the wave equation, there are both local and global components of the gradient in the descent flow that drives the surface evolution, with the added computational burden imposed by the translation of gradients between domains.

Regarding numerical aspects of research, general purpose acceleration techniques such as parallel and GPU processing may be justified at the developing stage of the wave equation assimilation algorithm if they significantly speed up the evaluation of the experiments

compared to the time and effort required to implement these techniques. A preconditioned Conjugate Gradient (PCG) algorithm may be a used as the numerical solver of a linearized version of the governing system of coupled PDEs of the model, with the multigrid methods developed in this thesis as the first choice of preconditioners.

Yet another interesting but more experimental research direction is the utilization of Principal Component Analysis (PCA) [71, 50] (also known as Proper Orthogonal Decomposition (POD) in physical contexts) and/or Dynamic Mode Decomposition (DMD) [83], as data analysis tools for the generation of predictive models that may be used to extract characteristic patterns to aid the surface reconstruction process.

APPENDIX A

NOTES ON CALCULUS OF VARIATIONS

In this appendix, we review basic results from Calculus of Variations that are used to derive governing equations of variational optimization problems. We limit our attention to results that involve first order derivatives of functionals.

To simplify notation, derivatives are denoted using subscripts. For example, the partial derivative with respect to x is written as $\partial_x(\cdot) \equiv \frac{\partial(\cdot)}{\partial x} \equiv (\cdot)_x$. In the case of a function that depends on a single variable, its derivative may also be written using Newton's notation: $\dot{x}(t) \equiv \frac{\mathrm{d}x}{\mathrm{d}t}$.

Since integration by parts is frequently used in the derivations, it is worth recalling it for future reference. Given two functions $a(x_1, \ldots, x_m)$ and $b(x_1, \ldots, x_m)$ for which the following integrals are meaningful, integration by parts states that:

$$\int_\Omega a_{x_i} b \, \mathrm{d}\Omega = -\int_\Omega a b_{x_i} \, \mathrm{d}\Omega + \int_{\partial\Omega} a b N^i \, \mathrm{d}\partial\Omega, \tag{117}$$

where N^i is the i-th component of the unit normal vector to the boundary $\partial\Omega$ at the considered point. The formula of integration by parts simplifies if we assume that the boundary integral vanishes:

$$\int_\Omega a_{x_i} b \, \mathrm{d}\Omega = -\int_\Omega a \, b_{x_i} \, \mathrm{d}\Omega. \tag{118}$$

Hence, it becomes a method to transfer derivatives between functions a and b.

A.1 Calculus of Variations. The Euler-Lagrange equations

Consider a functional $J : \mathcal{X} \to \mathbb{R}$ of the form

$$J(f) = \int_\Omega L(f, f_u, u) \, \mathrm{d}u, \tag{119}$$

where $L(f, f_u, u)$ is the Lagrangian of the functional J that maps a function $f \equiv f(u) \in \mathcal{X}$ into the real numbers and Ω is a fixed domain.

Suppose that one wishes to find the function f that minimizes (119). The Calculus of Variations (or Variational Calculus) is the branch of Mathematics that addresses this type of problems and it is the natural extension of the Multivariable (n-dimensional) Calculus to infinite dimensional spaces (e.g., the spaces of functions). The Euler-Lagrange (EL) equations state the necessary optimality condition that a function f must satisfy to be a minimizer of (130), and they are the equivalent condition for a *critical point* of a function in Multivariable Calculus.

Continuing with the analogy, in the same way that in Multivariable Calculus the directional derivative at the minimum of a function is zero for any possible perturbation direction, in Variational Calculus, the directional (Gâteaux) derivative at the minimum of a functional is zero. However, in the variational case, the direction is not defined by a vector but by a function. This idea will be used to derive the EL equations.

The directional (Gâteaux) derivative of J in the direction of the perturbation h (also called the *first variation* of J with respect to f) is

$$(D_h J)(f) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} J(f + \epsilon h)\bigg|_{\epsilon=0}. \tag{120}$$

Using the chain rule and integration by parts, it follows that

$$\begin{aligned}
(D_h J)(f) &= \frac{\mathrm{d}}{\mathrm{d}\epsilon} J(f + \epsilon h)\bigg|_{\epsilon=0} \\
&= \int_\Omega \frac{\mathrm{d}}{\mathrm{d}\epsilon} L(f + \epsilon h, f_u + \epsilon h_u, u)\bigg|_{\epsilon=0} \mathrm{d}u \\
&= \int_\Omega (L_f\, h + L_{f_u}\, h_u)\, \mathrm{d}u \\
&\overset{(117)}{=} \int_\Omega \big(hL_f - h(L_{f_u})_u\big)\, \mathrm{d}u + \int_{\partial\Omega} hL_{f_u}\, N^u\, \mathrm{d}\partial u.
\end{aligned}$$

Assuming that the space of admissible perturbations is such that $h|_{\partial\Omega} = 0$ for all h, the boundary term in the previous equation vanishes and we are left with

$$(D_h J)(f) = \int_\Omega (L_f - \frac{\partial}{\partial u} L_{f_u})\, h\, \mathrm{d}u. \tag{121}$$

For $f \in \mathcal{X}$ to be a minimizer of J, the directional derivative must be zero for all possible perturbations h, i.e. the first variation must vanish at f. Therefore, f must satisfy the EL

equations:

$$L_f - \frac{\partial}{\partial u} L_{f_u} = 0. \tag{122}$$

Introducing the L^2 inner product between two functions a and b:

$$\langle a, b \rangle_{L^2} = \int_\Omega a(u)\, b(u)\, \mathrm{d}u, \tag{123}$$

it is clear that the directional derivative (121) admits an inner product interpretation. If the gradient of the functional with respect to f is defined in terms of this inner product by

$$\nabla_f J \doteq L_f - \frac{\partial}{\partial u} L_{f_u}, \tag{124}$$

the directional derivative (121) becomes

$$(D_h J)(f) = \langle \nabla_f J, h \rangle_{L^2}, \tag{125}$$

which is a generalization of the formula in Multivariable Calculus of Euclidean space:

$$\frac{\partial g}{\partial \mathbf{v}} = \langle \nabla g, \mathbf{v} \rangle. \tag{126}$$

Finally, the EL equations (122) become $\nabla_f J = 0$, which is analogous to the equivalent condition in Multivariable Calculus ($\nabla g = 0$): at the optimal solution, the directional derivative must vanish regardless the direction. Observe that $\nabla_f J \equiv \nabla_f J(u)$ is a pointwise gradient, that is, it has a value for each $u \in \Omega$.

A.2 Gradient descent for functionals

The celebrated steepest descent optimization method can also be generalized from Multivariable Calculus to Variational Calculus. Therefore, it can be applied to minimize the functional (119). Assuming that the solution f also depends on an artificial time variable t, i.e., $f \equiv f(u, t)$, we set up a partial differential equation (PDE) of the form

$$\frac{\partial f}{\partial t} = -\nabla_f J(f(u, t)) \tag{127}$$

that is initialized by some guess $f(u, 0) = f_0(u)$. The steady state of the above PDE will satisfy

$$\frac{\partial f}{\partial t} = 0 \quad \Leftrightarrow \quad \nabla_f J(f(u, t)) = 0,$$

that is, the EL equations (122): the necessary optimality condition for a function f to be a minimizer of the functional $J(f)$.

The right side of (127) may be computed from the EL equations of the functional using (124) or from the time derivative of J due to the following result: the directional derivative (121) of J in the direction $h = f_t$ is equal to the time derivative of the energy functional:

$$J_t = (D_{f_t} J)(f). \tag{128}$$

This powerful link is easy to show: assuming u and t are independent, f is sufficiently smooth ($f \in C^2(\mathbb{R})$) to swap derivatives ($f_{ut} = f_{tu}$) and $h|_{\partial\Omega} = 0$:

$$\begin{aligned}
J_t &= \frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega L(f, f_u, u)\, \mathrm{d}u \\
&= \int_\Omega \frac{\mathrm{d}}{\mathrm{d}t} L(f, f_u, u)\, \mathrm{d}u \\
&= \int_\Omega \left(L_f\, f_t + L_{f_u}\, f_{ut}\right) \mathrm{d}u \\
&\overset{(117)}{=} \int_\Omega \left(L_f\, f_t - (L_{f_u})_u\, f_t\right) \mathrm{d}u + \int_{\partial\Omega} L_{f_u}\, f_t\, N^u\, \mathrm{d}\partial u \\
&\overset{(124)(125)}{=} \langle \nabla_f J, f_t \rangle_{L^2}
\end{aligned} \tag{129}$$

Observe, that by substituting (127) in (129) we obtain a negative quantity whose magnitude is the rate at which the functional decreases due to the choice of the perturbation direction f_t,

$$J_t = \langle \nabla_f J, \nabla_f J \rangle_{L^2} = -\|\nabla_f J\|_{L^2}^2 \leq 0.$$

Therefore, the family of functions $f(u, t)$ given by the evolution of (127) have decreasing value of J as the fictitious time parameter increases. Hence, the steady-state solution of (127) minimizes J. Moreover, by looking at (129) we see that the perturbation direction f_t given by (127) (or any other proportional one) is, locally, the most efficient perturbation that will decrease the value of J in time and hence the name "steepest descent".

A.3 Functionals of bi-parametric functions

The ideas in previous sections can be naturally extended for functionals of functions that depend on two variables. Consider a functional $J : \mathcal{X} \to \mathbb{R}$ of the form

$$J(f) = \int_\Omega L(f, f_u, f_v, u, v)\, \mathrm{d}u\mathrm{d}v, \tag{130}$$

where $L(f, f_u, f_v, u, v)$ is the Lagrangian of the functional J that maps a function $f \equiv f(u, v) \in \mathcal{X}$ into the real numbers and Ω is a fixed domain. The space of functions $\mathcal{X}$ is different than the one in previous sections, but we will use the same notation since they can be easily distinguished by the context.

The directional derivative of J still follows the general definition (120), but now the functions belong to a different space.

$$\begin{aligned}
(D_h J)(f) &\overset{(120)}{=} \int_\Omega \frac{\mathrm{d}}{\mathrm{d}\epsilon} L(f + \epsilon h, f_u + \epsilon h_u, f_v + \epsilon h_v, u, v)\Big|_{\epsilon=0} \mathrm{d}u\mathrm{d}v \\
&\overset{(117)}{=} \int_\Omega \big(hL_f - h(L_{f_u})_u - h(L_{f_v})_v\big)\, \mathrm{d}u\mathrm{d}v + \int_{\partial\Omega} h(L_{f_u} N^u + L_{f_v} N^v)\, \mathrm{d}\partial\Omega \\
&= \int_\Omega \big(L_f - \frac{\partial}{\partial u} L_{f_u} - \frac{\partial}{\partial v} L_{f_v}\big)\, h\, \mathrm{d}u\mathrm{d}v,
\end{aligned} \tag{131}$$

where we assumed a vanishing boundary term: $h|_{\partial\Omega} = 0,\ \forall h$. The EL equations are now

$$L_f - \frac{\partial}{\partial u} L_{f_u} - \frac{\partial}{\partial v} L_{f_v} = 0. \tag{132}$$

Following a parallel argument with respect to section A.1, we define the inner product between two functions a and b by

$$\langle a, b \rangle_{L^2} = \int_\Omega a(u, v)\, b(u, v)\, \mathrm{d}u\mathrm{d}v, \tag{133}$$

and define the gradient of the functional with respect to f accordingly by

$$\nabla_f J \doteq L_f - \frac{\partial}{\partial u} L_{f_u} - \frac{\partial}{\partial v} L_{f_v}, \tag{134}$$

so that formula (125) still holds (in the new function space) and (132) becomes $\nabla_f J = 0$.

The steepest descent equation is still (127), but using (134) on the right hand side since now $f \equiv f(u, v)$. Once more, the right hand side of (127) may be computed from the EL equations of the functional or from the time derivative of J since (129) still holds in the biparametric case.

APPENDIX B

ACTIVE SURFACES

The ideas presented in appendix A can also be extended to functionals of vector-valued functions, as it is the case of active contours and active surfaces. However, these use different (more geometric) inner products with respect to which the directional derivative and the gradient descent equations are referred. To understand the surface evolution presented in this thesis it is required a brief review of the ideas borrowed from Differential Geometry that are used in the calculation of the gradient descent flow that describes the surface evolution for the proposed energy functional.

B.1 Results from differential geometry of surfaces

Let u, v denote arbitrary parameters of a surface S, and let x, y, z denote coordinates of embedding space $\mathbb{R}^3$. Supose the surface can be covered by a single patch. A point on the surface can be described by

$$S(u,v) \equiv \mathbf{X}(u,v) = \big((u,v), y(u,v), z(u,v)\big)^\top. \tag{135}$$

The geometry of surfaces is richer than that of curves. In the following subsections we recall some results from the differential geometry of surfaces. Vectors $\mathbf{X}_u$ and $\mathbf{X}_v$ are the partial derivatives of the position vector (135). These vectors are tangent to the surface at the given point $\mathbf{X}$, i.e., they belong to the tangent space to the surface at a point $\mathbf{X}$, the so-called tangent plane $T_X(S)$.

The *First Fundamental Form* of a vector $\mathbf{w} = \mathbf{X}_u u' + \mathbf{X}_v v'$ in the tangent plane $T_X(S)$ to the surface S at $\mathbf{X}$, with coordinates $\mathbf{q} = (u', v')^\top$ with respect to the basis vectors, is the quadratic form

$$\mathbf{I}(\mathbf{w}) = \mathbf{w}\cdot\mathbf{w} = E(u')^2 + 2Fu'v' + G(v')^2 = (u', v') \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix} = \mathbf{q}^\top \mathcal{F}_I \mathbf{q}, \tag{136}$$

where $E = \mathbf{X}_u \cdot \mathbf{X}_u$, $F = \mathbf{X}_u \cdot \mathbf{X}_v$, $G = \mathbf{X}_v \cdot \mathbf{X}_v$ and $\mathcal{F}_I = \begin{pmatrix} E & F \\ F & G \end{pmatrix}$ is the matrix of coefficients of the First Fundamental Form.

The *area element* of the surface is given by

$$\mathrm{d}A = \|\mathbf{X}_u \times \mathbf{X}_v\| \, \mathrm{d}u\mathrm{d}v = \sqrt{EG - F^2} \, \mathrm{d}u\mathrm{d}v, \tag{137}$$

and the *unit normal* to the surface (at a point $\mathbf{X}$) in the outward direction is

$$\mathbf{N} = \frac{\mathbf{X}_u \times \mathbf{X}_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|} = \frac{\mathbf{X}_u \times \mathbf{X}_v}{\sqrt{EG - F^2}} \tag{138}$$

The *Second Fundamental Form* of a vector $\mathbf{w} = \mathbf{X}_u u' + \mathbf{X}_v v'$ in the tangent plane $T_X(S)$, with coordinates $\mathbf{q} = (u', v')^\top$ with respect to the basis vectors, is the quadratic form

$$\mathbf{II}(\mathbf{w}) = e(u')^2 + 2fu'v' + g(v')^2 = (u', v') \begin{pmatrix} e & f \\ f & g \end{pmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix} = \mathbf{q}^\top \mathcal{F}_{II} \mathbf{q}, \tag{139}$$

where $e = \mathbf{N} \cdot \mathbf{X}_{uu}$, $f = \mathbf{N} \cdot \mathbf{X}_{uv}$, $g = \mathbf{N} \cdot \mathbf{X}_{vv}$, and

$$\mathcal{F}_{II} = \begin{pmatrix} e & f \\ f & g \end{pmatrix} \tag{140}$$

is the matrix of coefficients of the Second Fundamental Form. The coefficients of the first and second fundamental forms satisfy certain compatibility conditions known as the Gauss-Codazzi equations; they involve the Christoffel symbols Γ_{ij}^k associated with the first fundamental form.

Weingarten's equations state that the partial derivatives of the unit normal to the surface, $\mathbf{N}_u$ and $\mathbf{N}_v$, are linear combinations of vectors $\mathbf{X}_u$ and $\mathbf{X}_v$. This is possible since all four vectors are in $T_X(S)$. Weingarten's equations in matrix form are

$$(\mathbf{N}_u, \mathbf{N}_v) = -(\mathbf{X}_u, \mathbf{X}_v)\mathcal{W}, \tag{141}$$

where $\mathcal{W}$ is the *Weingarten matrix* of the surface patch at $\mathbf{X}$ and it is related to the matrices of coefficients of the First and Second Fundamental Forms as follows:

$$\mathcal{W} = \mathcal{F}_I^{-1} \mathcal{F}_{II}. \tag{142}$$

The *Weingarten map* is also called the *shape operator*. It is the differential *df* of the Gauss map f and is a type of extrinsic curvature. The determinant and trace of the Weingarten matrix are related to the Gaussian curvature K and the mean curvature H of the surface at $\mathbf{X}$, respectively:

$$\det(\mathcal{W}) = \frac{eg - f^2}{EG - F^2} = K, \tag{143}$$

and

$$\text{trace}\,(\mathcal{W}) = 2H,$$

where

$$H = \frac{eG + gE - 2fF}{2(EG - F^2)}. \tag{144}$$

B.2 Gradient descent for functionals defined on surfaces

Consider the problem of finding the closed surface S that minimizes the energy functional defined by the weighted area:

$$J(S) = \int_S \Phi(\mathbf{X}, \mathbf{N})\,\mathrm{d}A, \tag{145}$$

where the weight Φ is a non-negative function and $\mathbf{N}$ is the outward normal to the surface.

Building upon ideas from section A.1 about Calculus of Variations, the surface S that minimizes (145) satisfies a set of EL equations. The gradient descent method yields a flow that can be used to find the minimizing surface. In the case of (145), the flow is, in general, a second order flow (see (156)) because it involves up to second order derivatives of Φ with respect to $\mathbf{X}$ and $\mathbf{N}$. We will use the two methods discussed in sections A.1 and A.2 to compute the right hand side of the gradient descent equation. The results therein require a fixed domain of integration, therefore, a change of integration variables is needed to express the problem in a suitable form: integration over the surface area is formulated as integration over a region in the parameter space $U \ni \mathbf{u} = (u, v)^\top$. In addition, note that the independent variable is a vector function, $S \equiv \mathbf{X}$, instead of a scalar function f.

Rewriting (145) in the standard form (130) (fixed domain of integration), yields

$$J = \int_U L(\mathbf{X}, \mathbf{X}_u, \mathbf{X}_v, u, v)\,\mathrm{d}u\mathrm{d}v, \tag{146}$$

where the Lagrangian is

$$L(\mathbf{X}, \mathbf{X}_u, \mathbf{X}_v, u, v) = \Phi(\mathbf{X}, \mathbf{N}) \, \|\mathbf{X}_u \times \mathbf{X}_v\|.$$

The EL equation of (146) is a vectorial version of (132):

$$L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v = \mathbf{0}. \tag{147}$$

Now, however, we wish to define $\nabla_S J$ (the gradient of the energy with respect to the surface) with respect to the area element instead of with respect to $\mathrm{d}u\mathrm{d}v$, to have a *geometric* (parameterization independent) definition. Therefore, the appropriate formula is

$$\nabla_S J \doteq \frac{L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|} \tag{148}$$

so that the directional derivative can be written as

$$\begin{aligned} (D_{\mathbf{w}} J)(S) &\overset{(131)}{=} \int_U \big(L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v\big) \cdot \mathbf{w} \, \mathrm{d}u\mathrm{d}v \\ &= \int_S \nabla_S J \cdot \mathbf{w} \, \mathrm{d}A. \end{aligned} \tag{149}$$

A vectorial version of (128) is also applicable:

$$J_t = (D_{S_t} J)(S). \tag{150}$$

This is easy to check following similar steps as those to arrive at (129): augmenting the surface with a fictitious time parameter t, makes the energy (145) dependent on such parameter and one may compute the time derivative: J_t. Finally, the gradient descent flow for (145) is the vectorial version of (127):

$$\mathbf{X}_t = -\nabla_S J. \tag{151}$$

There are two ways to compute $\nabla_S J$: (i) by computing the left hand side of the EL equation (147) and substituting in (148) or (ii) by computing the time derivative of J and manipulate the integral to express it as $\int_S \mathbf{B} \cdot \mathbf{X}_t \, \mathrm{d}A$, then concluding that $\nabla_S J = \mathbf{B}$.

B.2.1 Inner product interpretation

Assuming that a meaningful inner product can be defined in the space of closed surfaces by the formula

$$\langle \mathbf{U}, \mathbf{V} \rangle_{H^0} = \int_S \mathbf{U} \cdot \mathbf{V} \, \mathrm{d}A, \tag{152}$$

where $\mathbf{U}, \mathbf{V}$ are two vector fields defined on the surface S with area element $\mathrm{d}A$, and H^0 stands for the zero order Sobolev space (L^2), the directional derivative (149) becomes, according to (125),

$$(D_{\mathbf{w}} J)(S) = \langle \nabla_S J, \mathbf{w} \rangle_{H^0} \, , \tag{153}$$

and the time derivative (150) can be rewritten as the standard *geometric* H^0 inner product between two vector fields:

$$\frac{\mathrm{d}}{\mathrm{d}t} J(S) = \langle \nabla_S J, S_t \rangle_{H^0} \, . \tag{154}$$

B.3 Second order flow

Either using the time derivative approach or equation (148) from the EL approach, it can be shown that the gradient of the functional with respect to the surface is

$$\nabla_S J = \Phi_{\mathbf{X}} - \frac{1}{\|\mathbf{X}_u \times \mathbf{X}_v\|} \Big(\big(\mathbf{X}_v \times (\Phi \mathbf{N} + P_N^{\perp} \Phi_{\mathbf{N}}) \big)_u - \big(\mathbf{X}_u \times (\Phi \mathbf{N} + P_N^{\perp} \Phi_{\mathbf{N}}) \big)_v \Big), \tag{155}$$

where $P_{\mathbf{N}}^{\perp} = I - \mathbf{N}\mathbf{N}^{\top}$ is the projector onto $T_X(S)$. Carrying out the derivatives and simplifying, one can show that the gradient is a flow in the normal direction (as it should be for a geometric functional) involving up to second order derivatives of the integrand. The final expression is presented in [46]:

$$\nabla_S J = -\Big(2H \big(\Phi - (\Phi_{\mathbf{N}} \cdot \mathbf{N}) \big) - (\Phi_{\mathbf{X}} \cdot \mathbf{N}) - \operatorname{trace} (P_N^{\perp} \Phi_{\mathbf{NX}}) + \sum_{i=1}^{k} \lambda_i \mathbf{II}(P_N^{\perp} p_i) \Big) \, \mathbf{N}, \tag{156}$$

where $\mathbf{II}(\mathbf{w})$ is the Second Fundamental Form of the surface for a vector $\mathbf{w} \in T_X(S)$ and

$$\Phi_{\mathbf{NN}} = \sum_{i=1}^{k} \lambda_i p_i p_i^{\top}$$

is one of several valid decompositions of the symmetric matrix $\Phi_{\mathbf{NN}}$.

B.3.1 Special cases

There are a couple of special cases of the functional (145) that lead to simplified gradient descent flows, as it will be shown next.

Flow in divergence form. If the integrand of (145) admits a representation like $\Phi(\mathbf{X}, \mathbf{N}) = \mathbf{G}(\mathbf{X}) \cdot \mathbf{N}$, then $\Phi_{\mathbf{X}} = \mathbf{G}_{\mathbf{X}}\mathbf{N}$, $\Phi_{\mathbf{N}} = \mathbf{G}(\mathbf{X})$, $\Phi_{\mathbf{NX}} = \mathbf{G}_{\mathbf{X}}$, $\Phi_{\mathbf{NN}} = 0$ and the gradient of the functional (156) becomes

$$\nabla_S J = \big(\nabla \cdot \mathbf{G}\big)\,\mathbf{N} \tag{157}$$

Thus, the gradient descent flow (151) becomes first-order (no second derivatives of Φ appear in the final expression) because major simplifications occur. This is the generalization of the result for planar curves shown in [47], but applied to functionals on surfaces. For obvious reasons, the flow is said to be in *divergence form*. This situation naturally arises in many practical cases. In our case, it greatly simplifies the derivation of the formula of the variation of the data-fidelity energy with respect to the surface. As another example, it can also be used in [77] to simplify the involved calculations in the visibility maximization flows.

No dependence on the derivatives of the surface If the integrand $\Phi(\mathbf{X}, \mathbf{N}) = \Phi(\mathbf{X})$ does not depend on the derivatives of the surface, then all derivatives with respect to $\mathbf{N}$ vanish, yielding

$$\nabla_S J = -\big(2H\Phi - (\Phi_{\mathbf{X}} \cdot \mathbf{N})\big)\mathbf{N}.$$

In particular, if $\Phi = 1$, then $\Phi_{\mathbf{X}} = 0$ and the well known mean curvature flow ($\mathbf{X}_t = 2H\mathbf{N}$) is obtained, as expected.

APPENDIX C

CONSTRAINED GRAPH EVOLUTION

Here we establish the link between unconstrained surface evolution and surface evolution in the form of a graph, that is, in the case that the initial surface is represented as a graph and during the evolution it is forced to remain in the form of a graph, too. The main result states that the constrained evolution of the surface in the form of a graph is obtained by projecting the unconstrained evolution onto the direction of allowable variation of the surface as a graph. The theoretical connection will be exemplified with the simplest energy: Euclidean surface area.

C.1 Theoretical connection

Consider an energy functional $J(S)$ that depends on a surface S (and its derivatives) and can be written as (145)

$$J(S) = \int_S \Phi \, \mathrm{d}A = \int_U L(\mathbf{X}, \mathbf{X}_u, \mathbf{X}_v, u, v) \, \mathrm{d}u\mathrm{d}v.$$

The unconstrained gradient of the functional with respect to the surface, $\nabla_S J$, has already been introduced. Let us now consider the case of a surface that can be represented and evolved in the form of a graph, i.e., $\mathbf{X}(u, v) = (u, v, z(u, v))^\top$, $\mathbf{X}_u \times \mathbf{X}_v = (-z_u, -z_v, 1)^\top$ and $\mathbf{N} = \mathbf{X}_u \times \mathbf{X}_v / \|\mathbf{X}_u \times \mathbf{X}_v\|$. In this case, the functional becomes $\tilde{J}(z)$:

$$\tilde{J}(z) = \int_U \tilde{L}(z, z_u, z_v, u, v) \, \mathrm{d}u\mathrm{d}v, \tag{158}$$

or abusing notation, $J(z)$ with Lagrangian $L(z, z_u, z_v, u, v)$. Computing the first variation of the functional in this form with respect to z yields the classical result from Calculus of Variations (assuming that the boundary term resulting from integration by parts vanishes):

$$(D_w J)(z) = \int_U \big(L_z - (L_{z_u})_u - (L_{z_v})\big) \, w \, \mathrm{d}u\mathrm{d}v. \tag{159}$$

Now, let the gradient of the functional with respect to z, $\nabla_z J$, be measured with respect to the surface area element $\mathrm{d}A(u, v) = \|\mathbf{X}_u \times \mathbf{X}_v\| \, \mathrm{d}u\mathrm{d}v$, as in the case of the standard

geometric inner product for surfaces (152):

$$(D_w J)(z) = \int_U \nabla_z J \, w \, \mathrm{d}A(u, v), \tag{160}$$

Then, it follows that

$$\nabla_z J = \frac{L_z - (L_{z_u})_u - (L_{z_v})}{\|\mathbf{X}_u \times \mathbf{X}_v\|} = \frac{L_z - (L_{z_u})_u - (L_{z_v})}{\sqrt{1 + z_u^2 + z_v^2}}. \tag{161}$$

The rule that says how the graph of the surface must be evolved locally to have a maximum decrease of the energy functional is the gradient descent equation:

$$z_t = -\nabla_z J. \tag{162}$$

Is there a way to obtain the same results from the unconstrained gradient of the functional with respect to the surface, $\nabla_S J$? Yes, it is possible to enforce the graph constraint representation *after* computing the (unconstrained) gradient of the energy functional, as the following diagram suggests:

$$\begin{array}{ccc} J(S) & \xrightarrow{\text{graph}} & J(z) \\ \Big\downarrow{\scriptstyle \text{First variation}} & & \Big\downarrow \\ \nabla_S J & \overset{\text{projection}}{\dashrightarrow} & \nabla_z J \\ \Big\downarrow{\scriptstyle \text{Gradient descent}} & & \Big\downarrow \\ \mathbf{X}_t & \overset{\text{projection}}{\dashrightarrow} & z_t \end{array}$$

By the chain rule (see (204)) and (148),

$$\nabla_z J = \frac{\big(L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v\big) \cdot \mathbf{e}_3}{\|\mathbf{X}_u \times \mathbf{X}_v\|} = \nabla_S J \cdot \mathbf{e}_3. \tag{163}$$

That is, $\nabla_z J$ is the projection of the unconstrained gradient $\nabla_S J$ in the direction of allowable variation of the surface in the form of a graph, $\mathbf{e}_3 = (0, 0, 1)^\top$. This is very useful because in many practical cases it is easier to find an analytical formula for $\nabla_z J$ by means of (163) and exploiting existing results for $\nabla_S J$ (like (156) and (157)) rather than using (161).

The above link between gradients (163) yields the link between gradient descent PDEs for the unconstrained and the constrained cases:

$$z_t = -\nabla_z J = -\nabla_S J \cdot \mathbf{e}_3 = \mathbf{X}_t \cdot \mathbf{e}_3, \tag{164}$$

which is consistent with the time differentiation of $\mathbf{X}(u,v) = (u,v,z(u,v))^\top$ and the fact that $u_t = 0 = v_t$ since only the third component of $\mathbf{X}$ is allowed to vary in a graph representation: $\mathbf{X}_t = (0,0,z_t)^\top = z_t\mathbf{e}_3$. Finally, substituting (163) in (162) yields the descent equation in terms of the Lagrangian in $\mathbf{X}$ and its derivatives:

$$z_t = -\frac{L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|} \cdot \mathbf{e}_3. \tag{165}$$

Once more, it is clear that to compute the gradient descent equation for the graph only the projection of the left hand side of the E-L equations (147) in the direction of variation of the height is required (plus the speed factor $\|\mathbf{X}_u \times \mathbf{X}_v\|$).

Remarks. On the other hand, if the regular (non-geometric) L^2 inner product is used to define both gradients $\nabla_S J$ and $\nabla_z J$, e.g., by replacing $\mathrm{d}A(u,v)$ with $\mathrm{d}u\mathrm{d}v$ in both integrals $(D_{\mathbf{w}}J)(S)$ and $(D_w J)(z)$, equations $\nabla_z J = \nabla_S J \cdot \mathbf{e}_3$ and $z_t = \mathbf{X}_t \cdot \mathbf{e}_3$ will still hold, but with (148) replaced by $\nabla_S J = L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v$ (and consequently $\nabla_z J$ will also be altered). These are gradients in the L^2 sense, but they are not geometric or intrinsic to the surface because they are not parameterization-independent. The resulting gradient descent flows are not parameterization-independent, either.

Although the solution of the necessary optimality condition $(D_{\mathbf{w}}J)(S) = 0\ \forall \mathbf{w}$ (resp. $(D_w J)(z) = 0\ \forall w$) is unaffected by the non-zero factor $\|\mathbf{X}_u \times \mathbf{X}_v\|$, the evolution of the surface toward the solution is indeed affected by this speed factor.

For geometric functionals like the one considered in this section, it makes more sense to use a geometric inner product rather than a non-geometric one. For other functionals not arising from a geometric context it may be more convenient to consider the non-geometric inner product.

Alternative derivation. The above results can also be obtained using the fictitious time derivative approach. By augmenting the surface with a fictitious time variable in the form $\mathbf{X}(u,v,t)$, the functional also depends on t. Therefore, differentiating with respect to this parameter and after some effort, we arrive at (154). In the case of a graph, only the height of the surface is allowed to vary in time since the domain U remains fixed ($u_t = 0 = v_t$).

Hence,

$$\mathbf{X}(u,v,t) = (u,v,z(u,v,t))^\top,$$

and $\mathbf{X}_t = (0,0,z_t)^\top = z_t\mathbf{e}_3$. The latter implies (164). Substituting these equations in (154) and defining $\nabla_z J$ (the gradient of the energy with respect to z) in a geometric sense (because the surface area measure is used in the following integrals), yields the link shown in (163):

$$J_t = \langle \nabla_S J, z_t\mathbf{e}_3\rangle_{H^0} = \int_S \nabla_S J \cdot \mathbf{e}_3\, z_t \,\mathrm{d}A = \int_S \nabla_z J\, z_t \,\mathrm{d}A,$$

where

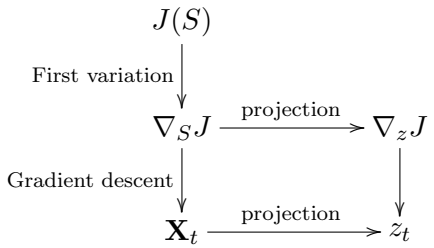
$$\nabla_z J \doteq \nabla_S J \cdot \mathbf{e}_3.$$

C.2 Example

Let the energy be the Euclidean area of the surface,

$$J(S) = \int_S \mathrm{d}A = \int_U \|\mathbf{X}_u \times \mathbf{X}_v\| \,\mathrm{d}u\mathrm{d}v = \int_U \sqrt{EG - F^2}\,\mathrm{d}u\mathrm{d}v. \tag{166}$$

C.2.1 *Method 1*: Enforcing the graph constraint after computing the variation of J with respect to the surface.



It is well known that the mean curvature flow arises in the variation of (166) with respect to the surface. In this case, $\nabla_S J = -2H\mathbf{N}$, and the flow that minimizes the area of a (closed) surface is (151):

$$\mathbf{X}_t = 2H\mathbf{N}. \tag{167}$$

For a surface described by an implicit function: $\phi(\mathbf{X}) = 0$ (e.g. a level set function), we further know the formula of the mean curvature in terms of the implicit function (see, for

example [87, p.70]):

$$2H = \nabla \cdot \frac{\nabla\phi}{\|\nabla\phi\|} = \frac{\left\{ \begin{array}{c} (\phi_{yy}+\phi_{zz})\phi_x^2 + (\phi_{zz}+\phi_{xx})\phi_y^2 + (\phi_{xx}+\phi_{yy})\phi_z^2 \\ -2\phi_x\phi_y\phi_{xy} - 2\phi_y\phi_z\phi_{yz} - 2\phi_z\phi_x\phi_{zx} \end{array} \right\}}{(\phi_x^2+\phi_y^2+\phi_z^2)^{3/2}}. \tag{168}$$

If we now impose the graph constraint on the surface (the initial surface is a graph and it remains in the form of a graph during the evolution on a fixed domain U), the corresponding descent equation is, by (164) and (167),

$$z_t = 2H\mathbf{N}\cdot\mathbf{e}_3 = 2H\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1} = 2H/\sqrt{1+z_u^2+z_v^2}. \tag{169}$$

Moreover, observing that the surface $Z = z(u,v)$ can be represented by the implicit function $\phi(u,v,Z) = Z - z(u,v)$, we have $\nabla\phi = (-z_u, -z_v, 1)^\top$ and we may obtain a formula for H in terms of the height of the graph, stemming from (168):

$$2H = \frac{(1+z_v^2)z_{uu} - 2z_u z_v z_{uv} + (1+z_u^2)z_{vv}}{(1+z_u^2+z_v^2)^{3/2}}. \tag{170}$$

Substituting (170) in (169), yields an integer power in the denominator of the right side:

$$z_t = \frac{(1+z_v^2)z_{uu} - 2z_u z_v z_{uv} + (1+z_u^2)z_{vv}}{(1+z_u^2+z_v^2)^{2}}.$$

Euler-Lagrange equations approach. The example energy (166) can be written in the standard form (146), whose Euler-Lagrange equations are (147). The gradient of the energy with respect to the surface is (148). If we now enforce the constrained surface variation in the form of a graph, (163), the resulting gradient descent equation for the graph is (165). Substituting the EL equations for energy (166) we conclude

$$z_t = \frac{2H(\mathbf{X}_u \times \mathbf{X}_v)}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\cdot\mathbf{e}_3 = 2H/\sqrt{1+z_u^2+z_v^2}. \tag{171}$$

C.2.2 *Method 2*: Enforcing the graph constraint before computing the variation of J with respect to the surface.

$$\begin{array}{ccc} J(S) & \xrightarrow{\text{graph}} & J(z) \\ & & \downarrow \text{First variation} \\ & & \nabla_z J \\ & & \downarrow \text{Gradient descent} \\ & & z_t \end{array}$$

Assuming that the surface is in the form of a graph, we may use this knowledge to rewrite the energy functional (166) as

$$J(S) = \int_U \sqrt{1 + z_u^2 + z_v^2}\,\mathrm{d}u\mathrm{d}v \equiv J(z). \tag{172}$$

Time derivative approach. Next, we augment the surface with a fictitious time variable in the form$\mathbf{X}(u, v, t) = (u, v, z(u, v, t))^\top$, that is, the height of the surface is allowed to vary in time, but the domain U remains fixed ($u_t = 0 = v_t$). Consequently, the functional also depends on t and differentiating with respect to this parameter, yields

$$\begin{aligned} J_t &= \int_U \big(\sqrt{1 + z_u^2 + z_v^2}\big)_t\,\mathrm{d}u\mathrm{d}v \\ &= \int_U \|\mathbf{X}_u \times \mathbf{X}_v\|^{-1}(z_u z_{tu} + z_v z_{tv})\,\mathrm{d}u\mathrm{d}v \\ &= \int_U -\Big(\big(\frac{z_u}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\big)_u + \big(\frac{z_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\big)_v\Big)\, z_t\,\mathrm{d}u\mathrm{d}v + \int_{\partial U} \cdots \\ &= \int_U (-2H\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1})\, z_t\,\mathrm{d}A(u, v), \end{aligned} \tag{173}$$

where we assumed a vanishing boundary term and we used

$$\begin{aligned} \big(\frac{z_u}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\big)_u + \big(\frac{z_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\big)_v &= \frac{z_{uu} + z_{vv}}{\|\mathbf{X}_u \times \mathbf{X}_v\|} - \frac{z_u(z_u z_{uu} + z_v z_{vu}) + z_v(z_u z_{uv} + z_v z_{vv})}{\|\mathbf{X}_u \times \mathbf{X}_v\|^3} \\ &= \frac{z_{uu} + z_{vv}}{\|\mathbf{X}_u \times \mathbf{X}_v\|} - \frac{z_u^2 z_{uu} + 2z_u z_v z_{vu} + z_v^2 z_{vv}}{\|\mathbf{X}_u \times \mathbf{X}_v\|^3} \\ &= \frac{1}{\|\mathbf{X}_u \times \mathbf{X}_v\|^3}\big((1 + z_v^2)z_{uu} - 2z_u z_v z_{uv} + (1 + z_u^2)z_{vv}\big) \\ &\overset{(170)}{=} 2H. \end{aligned}$$

From (160) and (173) it follows that

$$\nabla_z J = -2H\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1} = -2H/\sqrt{1 + z_u^2 + z_v^2} \tag{174}$$

and, by (162), the gradient descent equation to minimize (172) by varying the height z of the graph is (171).

Euler-Lagrange equations approach. Observe that the integrand of (172) is in the standard form (158). In this case, the Lagrangian L solely depends on the derivatives of z

(since $L_z = 0$), therefore, (161) becomes

$$\begin{aligned}
\nabla_z J &= -\big((L_{z_u})_u + (L_{z_v})_v\big)\,\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1} \\
&= -\Big(\big(\frac{z_u}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\big)_u + \big(\frac{z_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|}\big)_v\Big)\,\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1} \\
&= -2H\,\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1} \\
&= -2H/\sqrt{1 + z_u^2 + z_v^2},
\end{aligned}$$

which coincides with (174). Then, the descent equation (171) follows as before.

C.3 Application to the stereo surface reconstruction problem

Model 1. Consider the model (12), with energy components E_i (14), E_{rad} (20) and E_{geom} (19). If the gradients ∇_z and ∇_f are defined using the standard L^2 inner product, then, forgetting for a moment about the boundary conditions,

$$\begin{aligned}
\nabla_z E &= g(z) - \alpha\Delta z, \\
\nabla_f E &= -(\beta\Delta f - cf + K),
\end{aligned}$$

where $g(z) = \sum_i |\mathtt{M}^i|\tilde{Z}_i^{-3}(I_i - f)\big((\mathbf{X} - \mathbf{C}_i)\cdot\nabla\hat{f}\big)$, $c = \sum_i \mathtt{J}_i$ and $K = \sum_i I_i\mathtt{J}_i$. The corresponding gradient descent equations (in the parameter space U), are $z_t = -\nabla_z E$ and $f_t = -\nabla_f E$, that is,

$$\begin{aligned}
z_t &= \alpha\Delta z - g(z), \\
f_t &= \beta\Delta f - cf + K.
\end{aligned}$$

Model 2. However, if we keep E_i and E_{rad} in (12), but replace the regularizer in the shape of the surface by a more geometric choice: the surface area, $E_{\text{geom}} = \int_S \mathrm{d}A$, and at the same time the gradients ∇_z and ∇_f are defined using the weighted-area inner product (standard geometric L^2 inner product for surfaces), then, using a prime notation for the new gradients,

$$\begin{aligned}
\nabla'_z E &= \frac{g(z) - \alpha 2H}{\sqrt{1 + z_u^2 + z_v^2}}, \\
\nabla'_f E &= \frac{\nabla_f E}{\sqrt{1 + z_u^2 + z_v^2}},
\end{aligned}$$

and the corresponding gradient descent equations are $z_t = -\nabla'_z E$ and $f_t = -\nabla'_f E$, which include the extra factor $\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1} = (1 + z_u^2 + z_v^2)^{-1/2}$.

Discussion. As already mentioned, although the solution of the necessary optimality condition $(D_{(w,h)}E)(z, f) = 0 \; \forall (w, h)$ is unaffected by such a weight $\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1}$:

$$\begin{aligned} \nabla'_z E &= 0 = g(z) - \alpha 2H, \\ \nabla'_f E &= 0 = \nabla_f E, \end{aligned}$$

the evolution of the surface toward the optimal solution is indeed affected by this speed factor.

In practice, the gradient descent equation (time-stepping method) is only used to evolve the approximate solution of the height function z. The optimal radiance (for a fixed surface) is obtained by means of the Jacobi or the Gauss-Seidel (G-S) iteration because they are faster than the time-stepping method. The weight $\|\mathbf{X}_u \times \mathbf{X}_v\|^{-1}$ does not affect the Jacobi or G-S iteration since they are based on the discretization of the necessary optimality condition, not on the discretization of the gradient descent equation.

The two energy models proposed share the same radiance PDE, $\beta \Delta f - cf + K = 0$, but differ in the height PDE. This implies two distinct ways to evolve from an initial guess (z, f) toward the optimal solution (the one satisfying the necessary optimality condition for each model). The L^2 model dictates

$$z_t = \alpha \Delta z - g(z)$$

while the geometrically weighted (or simply called "geometric") model dictates

$$z_t = \frac{\alpha 2H - g(z)}{\sqrt{1 + z_u^2 + z_v^2}}.$$

The second model is a little more involved than the first one due to the expression for the mean curvature diffusive term and the square root in the denominator. However the square root only needs to be carried out for the non-linear term $g(z)$ since no square root appears in the diffusive term.

APPENDIX D

RECASTING IMAGE INTEGRALS

In this appendix we show how to recast integrals over the image domain to integrals over the surface or the parameter space. This is used in the data fidelity contribution of energy (12).

D.1 Jacobian of the change of variables

Let us derive an expression for the Jacobian of the change of integration variables from the image domain to the surface:

$$\mathtt{J}_i = \left| \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}} \right|. \tag{175}$$

Applying the chain rule to (17),

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}} = \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\tilde{\mathbf{X}}_i} \frac{\mathrm{d}\tilde{\mathbf{X}}_i}{\mathrm{d}\mathbf{X}} \frac{\mathrm{d}\mathbf{X}}{\mathrm{d}\mathbf{u}} = \frac{1}{\tilde{Z}_i^2} \begin{pmatrix} \tilde{Z}_i & 0 & -\tilde{X}_i \\ 0 & \tilde{Z}_i & -\tilde{Y}_i \end{pmatrix} \mathtt{M}^i \, (\mathbf{X}_u, \mathbf{X}_v). \tag{176}$$

Let $a_1^\top = (\tilde{Z}_i, 0, -\tilde{X}_i)$ and $a_2^\top = (0, \tilde{Z}_i, -\tilde{Y}_i)$, then

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}} = \frac{1}{\tilde{Z}_i^2} \begin{pmatrix} a_1^\top \\ a_2^\top \end{pmatrix} (\mathtt{M}^i \mathbf{X}_u, \mathtt{M}^i \mathbf{X}_v) = \frac{1}{\tilde{Z}_i^2} \begin{pmatrix} a_1^\top \mathtt{M}^i \mathbf{X}_u & a_1^\top \mathtt{M}^i \mathbf{X}_v \\ a_2^\top \mathtt{M}^i \mathbf{X}_u & a_2^\top \mathtt{M}^i \mathbf{X}_v \end{pmatrix},$$

whose determinant is

$$\det\left(\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}}\right) = \frac{1}{\tilde{Z}_i^4} \Big((a_1^\top \mathtt{M}^i \mathbf{X}_u)(a_2^\top \mathtt{M}^i \mathbf{X}_v) - (a_1^\top \mathtt{M}^i \mathbf{X}_v)(a_2^\top \mathtt{M}^i \mathbf{X}_u) \Big). \tag{177}$$

Let us simplify the right hand side of (177):

$$\begin{aligned}
&\tilde{Z}_i^{-4} \big((\mathtt{M}^i \mathbf{X}_u)^\top a_1 a_2^\top \mathtt{M}^i \mathbf{X}_v - (\mathtt{M}^i \mathbf{X}_v^\top) a_1 a_2^\top \mathtt{M}^i \mathbf{X}_v \big) \\
&= \tilde{Z}_i^{-4} (\mathtt{M}^i \mathbf{X}_u)^\top (a_1 a_2^\top - a_2 a_1^\top) \mathtt{M}^i \mathbf{X}_v \\
&= \tilde{Z}_i^{-4} (\mathtt{M}^i \mathbf{X}_u)^\top \big((a_2 \times a_1) \times (\mathtt{M}^i \mathbf{X}_v) \big) \\
&= \tilde{Z}_i^{-4} \det(\mathtt{M}^i \mathbf{X}_u, a_2 \times a_1, \mathtt{M}^i \mathbf{X}_v) \\
&= \tilde{Z}_i^{-4} \det(\mathtt{M}^i \mathbf{X}_u, -\tilde{Z}_i \tilde{\mathbf{X}}_i, \mathtt{M}^i \mathbf{X}_v) \\
&= \tilde{Z}_i^{-3} \det(\mathtt{M}^i) \det(\mathbf{X} + (\mathtt{M}^i)^{-1} \mathbf{p}_4^i, \mathbf{X}_u, \mathbf{X}_v) \\
&= \tilde{Z}_i^{-3} \det(\mathtt{M}^i) \det(\mathbf{X} - \mathbf{C}_i, \mathbf{X}_u, \mathbf{X}_v).
\end{aligned}$$

The last simplification used the fact that a point $\bar{\mathbf{C}}_i = (\mathbf{C}_i^\top, 1)^\top$ is the optical center of the i-th camera if it satisfies (2),

$$\mathtt{M}^i\mathbf{C}_i + \mathbf{p}_4^i = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{C}_i = -(\mathtt{M}^i)^{-1}\mathbf{p}_4^i. \tag{178}$$

Therefore, (177) becomes

$$\det\left(\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}}\right) = |\mathtt{M}^i|\tilde{Z}_i^{-3}(\mathbf{X} - \mathbf{C}_i)\cdot(\mathbf{X}_u \times \mathbf{X}_v) \tag{179}$$

where $\mathtt{M}^i = (\mathbf{n}_1^i, \mathbf{n}_2^i, \mathbf{n}_3^i)^\top$ is the left 3×3 sub-matrix of the projection matrix $\mathtt{P}^i = (p_{kl}^i)$, $|\mathtt{M}^i|$ is an alternative notation for $\det(\mathtt{M}^i)$, and $\tilde{Z}_i = \mathbf{n}_3^i \cdot \mathbf{X} + p_{34}^i$ can also be expressed as

$$\tilde{Z}_i = \mathbf{n}_3^i \cdot (\mathbf{X} - \mathbf{C}_i). \tag{180}$$

Here, $\tilde{Z}_i > 0$ is the depth of the point $\mathbf{X}$ with respect to the i-th camera (located at $\mathbf{C}_i$), as customary, in the direction of the normal $\mathbf{n}_3^i$ to the principal plane of the camera. We use the standard notation [35] that states that the depth is positive for points in front of the camera. Finally, assuming that the Jacobian is positive (the change of variables does not change the signs of oriented area elements between the surface and the image domain), it is the absolute value of (179): $\mathtt{J}_i = |\det(\,\mathrm{d}\mathbf{x}_i/\,\mathrm{d}\mathbf{u})|$ or (175), in standard notation. Thus,

$$\mathtt{J}_i = |\mathtt{M}^i|\tilde{Z}_i^{-3}\,|(\mathbf{X} - \mathbf{C}_i)\cdot(\mathbf{X}_u \times \mathbf{X}_v)|\,, \tag{181}$$

A necessary visibility condition. Equation (181) is not strictly correct since it does not take into account the visibility of the surface with respect to the camera. To account for this, recall that $\mathbf{X}_u \times \mathbf{X}_v$ is proportional to the outward unit normal to the surface at $\mathbf{X}(u, v)$ (138), and let us take into account the geometry of the camera setup with respect to the surface. Observe that $(\mathbf{X} - \mathbf{C}_i)\cdot\mathbf{N} < 0$ for neighborhoods of surface points (i.e., patches) pointing toward the camera and $(\mathbf{X} - \mathbf{C}_i)\cdot\mathbf{N} > 0$ for patches pointing away from the camera. The latter are occluded by the former from the viewpoint of the camera. Hence,

$$\mathtt{J}_i = |\mathtt{M}^i|\tilde{Z}_i^{-3}\max\big(-(\mathbf{X} - \mathbf{C}_i)\cdot(\mathbf{X}_u \times \mathbf{X}_v), 0\big). \tag{182}$$

Beware that, for a given surface point $\mathbf{X}$, the condition of positive Jacobian is not sufficient for that point to be visible from the camera viewpoint since the surface may be

self-occluded. Therefore, a positive Jacobian is a necessary visibility condition, but not a sufficient condition.

D.2 Area measures in the image and on the surface

With the expression of the Jacobian of the change of variables at hand (182), it is straightforward to derive a formula for the relationship between area elements in the image plane and on the surface:

$$\mathrm{d}\mathbf{x}_i = \mathrm{d}x_i\,\mathrm{d}y_i = \left|\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}}\right| \mathrm{d}\mathbf{u} = \mathtt{J}_i\ \mathrm{d}\mathbf{u} = |\mathtt{M}^i|\tilde{Z}_i^{-3}\max\big(-(\mathbf{X}-\mathbf{C}_i)\cdot(\mathbf{X}_u\times\mathbf{X}_v),0\big)\,\mathrm{d}u\mathrm{d}v. \quad (183)$$

Since the surface area element is (137) and the outward unit normal to the surface at $\mathbf{X}(u,v)$ is (138), the relationship between area elements (183) can be rewritten as

$$\mathrm{d}\mathbf{x}_i = |\mathtt{M}^i|\tilde{Z}_i^{-3}\max\big(-(\mathbf{X}-\mathbf{C}_i)\cdot\mathbf{N},0\big)\,\mathrm{d}A. \quad (184)$$

The term $(\mathbf{X}-\mathbf{C}_i)\cdot\mathbf{N}$ is proportional to the cosine of the angle between the unit normal to the surface at $\mathbf{X}$ and the ray joining the optical center of the camera and $\mathbf{X}$, i.e., the *projection ray*. One may observe the extreme cases:

- If $(\mathbf{X}-\mathbf{C}_i)\perp\mathbf{N}$, the surface patch at $\mathbf{X}$ projects to a line in the image plane, hence $d\mathbf{x}_i = 0$ (zero area) and that patch makes no contribution to the energy E_i.
- On the other hand, if the ray from the camera to the surface point $\mathbf{X}$ is parallel to the normal of the surface patch at that point, i.e., $(\mathbf{X}-\mathbf{C}_i)\parallel\mathbf{N}$, the surface patch projects onto a maximum area region $d\mathbf{x}_i$ (assuming the patch is facing the camera).

This qualitative behavior of the model agrees with our physical intuition.

In the next sections, to simplify calculations related to the evolution of the surface height and radiance according to the data fidelity term we will use the former expression for the Jacobian (181) that does not take into account the necessary visibility condition, i.e.

$$\mathtt{J}_i = -|\mathtt{M}^i|\tilde{Z}_i^{-3}(\mathbf{X}-\mathbf{C}_i)\cdot(\mathbf{X}_u\times\mathbf{X}_v). \quad (185)$$

However, keep in mind that if the surface point under consideration is not visible, it will not will be allowed to evolve according to the data fidelity flow derived.

APPENDIX E

THE EULER-LAGRANGE EQUATIONS

In this section it will be shown how to calculate the necessary optimality conditions to minimize the proposed energy functional (12). The variation of the energy with respect to the surface radiance will be presented first because it is easier to compute than the variation with respect to the shape.

E.1 Variation with respect to the surface radiance

Let us derive the PDE for the first variation of the energy with respect to the radiance (35). Since E_{geom} does not depend on the radiance f, it has no effect on the aforementioned first variation. Straightforward calculations show that for the simple regularizer (29),

$$(L_{\text{rad}})_f - \big((L_{\text{rad}})_{f_u}\big)_u - \big((L_{\text{rad}})_{f_v}\big)_v = -f_{uu} - f_{vv} = -\Delta f. \tag{186}$$

Focusing now on the data fidelity term, L_i does not depend on the gradient of f. Therefore

$$(L_i)_f - \big((L_i)_{f_u}\big) - \big((L_i)_{f_v}\big) = (\phi_i)_f \mathtt{J}_i = -(I_i - f)\mathtt{J}_i. \tag{187}$$

Assuming a setup with N_c images of the same scene from different viewpoints, it is straightforward to derive (39) by substituting (186), (187), and (25) in (35) and applying linearity. Observe that if $\beta = 0$, the resulting PDE (35) is linear in the unknown f and admits a closed form solution: the optimal f is the weighted average

$$f = \sum_{i=1}^{N_c} w_i I_i, \qquad w_i = \frac{\mathtt{J}_i}{\sum_{j=1}^{N_c} \mathtt{J}_j}, \tag{188}$$

where the weights w_i may not yield a convex combination because the non-negative Jacobians might all vanish for an occluded surface point.

E.1.1 Boundary condition for the PDE in the radiance of the surface

The PDE (35) comes with natural homogeneous Neumann boundary conditions, as it will be shown. The regularizer (29) yields the directional derivative of f along $\boldsymbol{\nu}$, the unit normal

to the integration domain U in the parameter space.

$$(L_{\text{rad}})_{f_u}\nu^u + (L_{\text{rad}})_{f_v}\nu^v = f_u\nu^u + f_v\nu^v = \frac{\partial f}{\partial \boldsymbol{\nu}}. \tag{189}$$

Because L_i and L_{geom} do not depend on the gradient of f, the left hand side of (36) is

$$L_{f_u}\nu^u + L_{f_v}\nu^v = \beta\frac{\partial f}{\partial \boldsymbol{\nu}}. \tag{190}$$

If $\beta \neq 0$, it follows that (36) is equivalent to the Neumann boundary condition, on ∂U,

$$\frac{\partial f}{\partial \boldsymbol{\nu}} = 0. \tag{191}$$

E.2 3-D extensions of the radiance and the images

To simplify the calculations involved in the EL equations, let us define extensions of the radiance and image intensities to the whole three-dimensional space, $\hat{f} : \mathbb{R}^3 \to \mathbb{R}$ and $\hat{I}_i : \mathbb{R}^3 \to \mathbb{R}$, respectively. It is natural to define the latter as being constant along optical rays (projection rays) from the camera,

$$\hat{I}_i\Big|_S \equiv \hat{I}_i(\mathbf{X}) \doteq I_i\big(\pi_i(\mathbf{X})\big). \tag{192}$$

There is a degree of freedom in the specification of the extension of the radiance. One could choose $\hat{f}$ to be constant along the normal to the surface. In this case, the values of f would be propagated along the normal to the surface. However, since we will be using this extension for surfaces in graph form (16), let us define the extension to be constant along the third dimension, i.e., the Z axis, as mentioned in section 3.2:

$$\hat{f}\Big|_S \equiv \hat{f}\big(\mathbf{X}(\mathbf{u})\big) \doteq f(\mathbf{u}). \tag{193}$$

In a world coordinate system adapted to the graph, where the parameter space of the surface is the plane $Z = 0$, equation (193) is equivalent to

$$\hat{f}(X, Y, Z) \doteq f(X, Y). \tag{194}$$

From (192) and (193), the photometric matching criterion (15) can also be extended to the whole space, $\hat{\phi}_i : \mathbb{R}^3 \to \mathbb{R}$, by the definition:

$$\hat{\phi}_i(\mathbf{X}) \doteq \tfrac{1}{2}\big(\hat{I}_i(\mathbf{X}) - \hat{f}(\mathbf{X})\big)^2. \tag{195}$$

It is clear that for surface points the restriction of the extension satisfies $\hat{\phi}_i|_S \equiv \phi_i$.

E.3 Variation with respect to the shape of the surface

Now, let us compute each term in the left hand side of (33). Since L_{rad} does not depend on Z, it has no effect on (33). On the other hand, straightforward calculations show that the chosen simple regularizer (28) yields, as in (186),

$$(L_{\text{geom}})_Z - \big((L_{\text{geom}})_{Z_u}\big)_u - \big((L_{\text{geom}})_{Z_v}\big)_v = -Z_{uu} - Z_{vv} = -\Delta Z. \tag{196}$$

Let us focus now on the data fidelity term. The extensions defined in section E.2 make it possible to rewrite the Lagrangian L_i in (27) as a function of $\mathbf{X}$,

$$L_i = \hat{\phi}_i(\mathbf{X}, \hat{f})\mathtt{J}_i(\mathbf{X}, \mathbf{X}_u, \mathbf{X}_v). \tag{197}$$

The chain rule can be used to compute the left hand side of (33) for L_i because the derivatives in Z are a projection of the derivatives in X:

$$L_Z = L_{\mathbf{X}} \cdot \mathbf{X}_Z = L_{\mathbf{X}} \cdot \mathbf{e}_3, \tag{198}$$

$$L_{Z_u} = L_{\mathbf{X}_u} \cdot (\mathbf{X}_u)_{Z_u} + L_{\mathbf{X}_v} \cdot (\mathbf{X}_v)_{Z_u} = L_{\mathbf{X}_u} \cdot \mathbf{e}_3, \tag{199}$$

$$L_{Z_v} = L_{\mathbf{X}_u} \cdot (\mathbf{X}_u)_{Z_v} + L_{\mathbf{X}_v} \cdot (\mathbf{X}_v)_{Z_v} = L_{\mathbf{X}_v} \cdot \mathbf{e}_3, \tag{200}$$

where $\mathbf{e}_3 = (0, 0, 1)^\top$ is the direction of variation of the height. The derivatives of $\mathbf{X}(u, v)$ given by (16) are

$$\mathbf{X}_u = (1, 0, Z_u)^\top, \tag{201}$$

$$\mathbf{X}_v = (0, 1, Z_v)^\top, \tag{202}$$

$$\mathbf{X}_u \times \mathbf{X}_v = (-Z_u, -Z_v, 1)^\top. \tag{203}$$

That is,

$$L_Z - (L_{Z_u})_u - (L_{Z_v})_v = \big(L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v\big) \cdot \mathbf{e}_3. \tag{204}$$

Therefore, the left hand side of (33) for L_i is

$$(L_i)_Z - \big((L_i)_{Z_u}\big)_u - \big((L_i)_{Z_v}\big)_v = \Big((L_i)_{\mathbf{X}} - \big((L_i)_{\mathbf{X}_u}\big)_u - \big((L_i)_{\mathbf{X}_v}\big)_v\Big) \cdot \mathbf{e}_3. \tag{205}$$

Now, it remains to calculate $(L_i)_{\mathbf{X}}$, $\big((L_i)_{\mathbf{X}_u}\big)_u$ and $\big((L_i)_{\mathbf{X}_v}\big)_v$. The first term is

$$(L_i)_{\mathbf{X}} = (\hat{\phi}_i \mathtt{J}_i)_{\mathbf{X}} = (\hat{\phi}_i)_{\mathbf{X}} \mathtt{J}_i + \hat{\phi}_i (\mathtt{J}_i)_{\mathbf{X}}, \tag{206}$$

where

$$(\hat{\phi}_i)_{\mathbf{X}} = (\hat{I}_i - \hat{f})(\nabla \hat{I}_i - \nabla \hat{f}), \tag{207}$$

$$(\mathtt{J}_i)_{\mathbf{X}} = -|\mathtt{M}^i|\Big(-3\tilde{Z}_i^{-4}\big((\mathbf{X}-\mathbf{C}_i)\cdot(\mathbf{X}_u\times\mathbf{X}_v)\big)\mathbf{n}_3^i + \tilde{Z}_i^{-3}(\mathbf{X}_u\times\mathbf{X}_v)\Big). \tag{208}$$

As is customary, let ∇ denote the spatial derivative, then image derivatives will arise in the calculations regarding (207). By the chain rule,

$$\nabla \hat{I}_i^\top = \nabla I_i^\top \frac{\partial \pi_i}{\partial \mathbf{X}}, \tag{209}$$

with $\partial \pi_i / \partial \mathbf{X}$ as in (176):

$$\frac{\partial \pi_i}{\partial \mathbf{X}} \equiv \frac{\partial \mathbf{x}_i}{\partial \tilde{\mathbf{X}}^i}\frac{\partial \tilde{\mathbf{X}}^i}{\partial \mathbf{X}}, \tag{210}$$

where the derivative of the image coordinates with respect to the camera coordinates is

$$\frac{\partial \mathbf{x}_i}{\partial \tilde{\mathbf{X}}^i} = \frac{1}{\tilde{Z}_i^2}\begin{pmatrix} \tilde{Z}_i & 0 & -\tilde{X}_i \\ 0 & \tilde{Z}_i & -\tilde{Y}_i \end{pmatrix}. \tag{211}$$

and the derivative of the camera coordinates with respect to the world coordinates is

$$\frac{\partial \tilde{\mathbf{X}}^i}{\partial \mathbf{X}} = \mathtt{M}^i. \tag{212}$$

As a space point $\mathbf{X}$ moves along the optical ray from a camera, the corresponding image point $\mathbf{x}_i = \pi_i(\mathbf{X})$ remains unchanged. This implies the following remarkable fact:

$$\frac{\partial \pi_i}{\partial \mathbf{X}}(\mathbf{X} - \mathbf{C}_i) = \mathbf{0}. \tag{213}$$

Proof. Using the chain rule, equations (178) and (212), it yields

$$\frac{\partial \pi_i}{\partial \mathbf{X}}(\mathbf{X}-\mathbf{C}_i) = \frac{\partial \mathbf{x}_i}{\partial \tilde{\mathbf{X}}^i}\frac{\partial \tilde{\mathbf{X}}^i}{\partial \mathbf{X}}(\mathbf{X}-\mathbf{C}_i) = \frac{\partial \mathbf{x}_i}{\partial \tilde{\mathbf{X}}^i}\mathtt{M}^i(\mathbf{X}-\mathbf{C}_i) = \frac{\partial \mathbf{x}_i}{\partial \tilde{\mathbf{X}}^i}(\mathtt{M}^i\mathbf{X} + \mathbf{p}_4^i) = \frac{\partial \mathbf{x}_i}{\partial \tilde{\mathbf{X}}^i}\tilde{\mathbf{X}}^i = \mathbf{0}. \ \square$$

Combining (209) and (213) one can show that, since the intensity of the extension $\hat{I}_i$ is constant along the projection ray, $\nabla \hat{I}_i$ lies in the plane orthogonal to such projection ray:

$$\nabla \hat{I}_i \cdot (\mathbf{X} - \mathbf{C}_i) = 0. \tag{214}$$

This result will lead to a useful simplification of the term $(\hat{\phi}_i)_{\mathbf{X}}\mathtt{J}_i$ in (206) that will have an important consequence in the final result: no derivatives of the image data appear in the

final EL equations. This desirable feature makes the algorithm less sensitive to image noise when compared to other variational approaches for stereo 3-D reconstruction. This feature is shared by the standard Mumford-Shah [66] formulation for direct image segmentation. In our case, it arises from the fact that the stereo discrepancy is measured in the image domain rather than on the surface [93]. Observe that it is a purely geometric result, thus independent of the choice of ϕ_i.

Now, let us turn our attention to the second and third terms in (205). Since $\hat{\phi}_i$ in (195) does not depend on the normal to the surface, it follows that $(\hat{\phi}_i)_{\mathbf{X}_u} = 0$ and $(\hat{\phi}_i)_{\mathbf{X}_v} = 0$. Therefore,

$$(L_i)_{\mathbf{X}_u} = (\hat{\phi}_i \mathtt{J}_i)_{\mathbf{X}_u} = (\hat{\phi}_i)_{\mathbf{X}_u} \mathtt{J}_i + \hat{\phi}_i (\mathtt{J}_i)_{\mathbf{X}_u} = \hat{\phi}_i (\mathtt{J}_i)_{\mathbf{X}_u}, \tag{215}$$

where

$$\begin{aligned} (\mathtt{J}_i)_{\mathbf{X}_u} &= \Big(-|\mathtt{M}^i| \tilde{Z}_i^{-3} \big((\mathbf{X} - \mathbf{C}_i) \cdot (\mathbf{X}_u \times \mathbf{X}_v) \big) \Big)_{\mathbf{X}_u} \\ &= -|\mathtt{M}^i| \tilde{Z}_i^{-3} \Big(\mathbf{X}_u \cdot \big(\mathbf{X}_v \times (\mathbf{X} - \mathbf{C}_i) \big) \Big)_{\mathbf{X}_u} \\ &= -|\mathtt{M}^i| \tilde{Z}_i^{-3} \big(\mathbf{X}_v \times (\mathbf{X} - \mathbf{C}_i) \big). \end{aligned} \tag{216}$$

Similarly,

$$(L_i)_{\mathbf{X}_v} = (\hat{\phi}_i \mathtt{J}_i)_{\mathbf{X}_v} = (\hat{\phi}_i)_{\mathbf{X}_v} \mathtt{J}_i + \hat{\phi}_i (\mathtt{J}_i)_{\mathbf{X}_v} = \hat{\phi}_i (\mathtt{J}_i)_{\mathbf{X}_v} \tag{217}$$

with

$$\begin{aligned} (\mathtt{J}_i)_{\mathbf{X}_v} &= \Big(-|\mathtt{M}^i| \tilde{Z}_i^{-3} \big((\mathbf{X} - \mathbf{C}_i) \cdot (\mathbf{X}_u \times \mathbf{X}_v) \big) \Big)_{\mathbf{X}_v} \\ &= -|\mathtt{M}^i| \tilde{Z}_i^{-3} \Big(\mathbf{X}_v \cdot \big((\mathbf{X} - \mathbf{C}_i) \times \mathbf{X}_u \big) \Big)_{\mathbf{X}_v} \\ &= -|\mathtt{M}^i| \tilde{Z}_i^{-3} \big((\mathbf{X} - \mathbf{C}_i) \times \mathbf{X}_u \big). \end{aligned} \tag{218}$$

Differentiating (215) with respect to u gives

$$\big((L_i)_{\mathbf{X}_u} \big)_u = \big(\hat{\phi}_i (\mathtt{J}_i)_{\mathbf{X}_u} \big)_u = (\hat{\phi}_i)_u (\mathtt{J}_i)_{\mathbf{X}_u} + \hat{\phi}_i \big((\mathtt{J}_i)_{\mathbf{X}_u} \big)_u, \tag{219}$$

where

$$(\hat{\phi}_i)_u = (\hat{I}_i - \hat{f})(\hat{I}_i - \hat{f})_u = (\hat{I}_i - \hat{f})(\nabla \hat{I}_i - \nabla \hat{f}) \cdot \mathbf{X}_u, \tag{220}$$

$$\big((\mathtt{J}_i)_{\mathbf{X}_u} \big)_u = -|\mathtt{M}^i| \Big(-3 \tilde{Z}_i^{-4} (\mathbf{n}_3^i \cdot \mathbf{X}_u) \big(\mathbf{X}_v \times (\mathbf{X} - \mathbf{C}_i) \big) + \tilde{Z}_i^{-3} \big(\mathbf{X}_{vu} \times (\mathbf{X} - \mathbf{C}_i) + \mathbf{X}_v \times \mathbf{X}_u \big) \Big). \tag{221}$$

Similarly, differentiating (217) with respect to v yields

$$\big((L_i)_{\mathbf{X}_v}\big)_v = \big(\hat{\phi}_i(\mathtt{J}_i)_{\mathbf{X}_v}\big)_v = (\hat{\phi}_i)_v(\mathtt{J}_i)_{\mathbf{X}_v} + \hat{\phi}_i\big((\mathtt{J}_i)_{\mathbf{X}_v}\big)_v, \tag{222}$$

where

$$(\hat{\phi}_i)_v = (\hat{I}_i - \hat{f})(\hat{I}_i - \hat{f})_v = (\hat{I}_i - \hat{f})(\nabla\hat{I}_i - \nabla\hat{f}) \cdot \mathbf{X}_v, \tag{223}$$

$$\big((\mathtt{J}_i)_{\mathbf{X}_v}\big)_v = -|\mathtt{M}^i|\Big(-3\tilde{Z}_i^{-4}(\mathbf{n}_3^i\cdot\mathbf{X}_v)\big((\mathbf{X}-\mathbf{C}_i)\times\mathbf{X}_u\big)+\tilde{Z}_i^{-3}\big((\mathbf{X}-\mathbf{C}_i)\times\mathbf{X}_{uv}+\mathbf{X}_v\times\mathbf{X}_u\big)\Big). \tag{224}$$

Next, assume the surface is sufficiently smooth such that it is twice continuously differentiable. Therefore, the second order mixed derivatives are equal, $\mathbf{X}_{uv} = \mathbf{X}_{vu}$. Substituting previous partial results (219)-(224) in (205) yields

$$\begin{aligned}
&(L_i)_{\mathbf{X}} - \big((L_i)_{\mathbf{X}_u}\big)_u - \big((L_i)_{\mathbf{X}_v}\big)_v \\
&= (\hat{\phi}_i)_{\mathbf{X}}\mathtt{J}_i - (\hat{\phi}_i)_u(\mathtt{J}_i)_{\mathbf{X}_u} - (\hat{\phi}_i)_v(\mathtt{J}_i)_{\mathbf{X}_v} + \hat{\phi}_i\Big((\mathtt{J}_i)_{\mathbf{X}} - \big((\mathtt{J}_i)_{\mathbf{X}_u}\big)_u - \big((\mathtt{J}_i)_{\mathbf{X}_v}\big)_v\Big), \\
&= -|\mathtt{M}^i|\tilde{Z}_i^{-3}\Big((\hat{I}_i - \hat{f})\,\mathbf{W}(\nabla\hat{I}_i - \nabla\hat{f}) + 3\hat{\phi}_i\big((\mathbf{X}_u\times\mathbf{X}_v) - \tilde{Z}_i^{-1}\mathbf{W}(\mathbf{n}_3^i)\big)\Big),
\end{aligned} \tag{225}$$

where we define the vector

$$\mathbf{W}(\mathbf{b}) = -(\mathbf{X}-\mathbf{C}_i)\cdot(\mathbf{X}_u\times\mathbf{X}_v)\,\mathbf{b}+(\mathbf{b}\cdot\mathbf{X}_u)\big(\mathbf{X}_v\times(\mathbf{X}-\mathbf{C}_i)\big)+(\mathbf{b}\cdot\mathbf{X}_v)\big((\mathbf{X}-\mathbf{C}_i)\times\mathbf{X}_u\big). \tag{226}$$

Next, we show that this vector is proportional to the unit normal $\mathbf{N}$ and the value of $\mathbf{b}$ only affects its magnitude. Let $\mathtt{A} = (\mathbf{X} - \mathbf{C}_i, \mathbf{X}_u, \mathbf{X}_v)$, then

$$\det(\mathtt{A})\mathtt{A}^{-\top} = \big(\mathbf{X}_u \times \mathbf{X}_v,\ \mathbf{X}_v \times (\mathbf{X} - \mathbf{C}_i),\ (\mathbf{X} - \mathbf{C}_i) \times \mathbf{X}_u\big), \tag{227}$$

and, from $\mathtt{I} = \mathtt{A}^{-\top}\mathtt{A}^{\top}$, we derive the following matrix relation:

$$\det(\mathtt{A})\mathtt{I} - \big(\mathbf{X}_v \times (\mathbf{X} - \mathbf{C}_i)\big)\mathbf{X}_u^{\top} - \big((\mathbf{X} - \mathbf{C}_i) \times \mathbf{X}_u\big)\mathbf{X}_v^{\top} = (\mathbf{X}_u \times \mathbf{X}_v)(\mathbf{X} - \mathbf{C}_i)^{\top}, \tag{228}$$

which can be used in (226) to obtain

$$\mathbf{W}(\mathbf{b}) = (\mathbf{X}_u \times \mathbf{X}_v)(\mathbf{X} - \mathbf{C}_i)^{\top}\mathbf{b}. \tag{229}$$

Substituting (229) in (225), and using the alternative definition of the depth $\tilde{Z}_i$ (180) and (214), yields important simplifications: the term that is multiplying $\hat{\phi}_i$ vanishes and no

image derivatives appear in the final expression.

$$\begin{aligned}
&(L_i)_{\mathbf{X}} - \big((L_i)_{\mathbf{X}_u}\big)_u - \big((L_i)_{\mathbf{X}_v}\big)_v \\
&= |\mathtt{M}^i|\tilde{Z}_i^{-3}\Big((\hat{I}_i - \hat{f})\big((\mathbf{X} - \mathbf{C}_i)\cdot\nabla\hat{f}\big) - 3\hat{\phi}_i(1 - \tilde{Z}_i^{-1}\tilde{Z}_i)\Big)(\mathbf{X}_u \times \mathbf{X}_v) \\
&= |\mathtt{M}^i|\tilde{Z}_i^{-3}(\hat{I}_i - \hat{f})\big((\mathbf{X} - \mathbf{C}_i)\cdot\nabla\hat{f}\big)(\mathbf{X}_u \times \mathbf{X}_v).
\end{aligned} \tag{230}$$

Therefore, according to $(\mathbf{X}_u \times \mathbf{X}_v)\cdot\mathbf{e}_3 = 1$ (see (203)), the left hand side of (205) becomes

$$(L_i)_Z - \big((L_i)_{Z_u}\big)_u - \big((L_i)_{Z_v}\big)_v = |\mathtt{M}^i|\tilde{Z}_i^{-3}(\hat{I}_i - \hat{f})\big((\mathbf{X} - \mathbf{C}_i)\cdot\nabla\hat{f}\big). \tag{231}$$

The freedom in the definition of the extension $\hat{f}$ (193) allows for further simplification of (231) (in the coordinate system adapted to the graph representation of the surface). The chosen definition yields $\nabla\hat{f} = \big((\nabla f)^\top, 0\big)^\top$, which will cancel the third coordinate of $(\mathbf{X} - \mathbf{C}_i)$, leaving the EL equation in terms of the original images I_i, the radiance function f and the optical center of the i-th camera $\mathbf{C}_i = (C_i^1, C_i^2, C_i^3)^\top$,

$$(L_i)_Z - \big((L_i)_{Z_u}\big)_u - \big((L_i)_{Z_v}\big)_v = |\mathtt{M}^i|\tilde{Z}_i^{-3}(I_i - f)(u - C_i^1, v - C_i^2)^\top\nabla f. \tag{232}$$

The terms indirectly affected by the height Z are the image intensity at the current 3-D point, $I_i \equiv I_i(\mathbf{x}_i(\mathbf{X}(u, v, Z(u, v))))$, and the depth of the 3-D point with respect to the camera, $\tilde{Z}_i = \mathbf{n}_3^i \cdot \mathbf{X}(u, v, Z(u, v)) + p_{34}^i$.

Assuming a setup with N_c images of the same scene from different viewpoints, one can directly derive the EL equation for the variation with respect to the shape of the surface of the total energy since all integrals are expressed with respect to the same domain (the parameterization plane). Hence, collecting terms (196) and (232), PDE (33) becomes (37):

$$\Big(\sum_{i=1}^{N_c} |\mathtt{M}^i|\tilde{Z}_i^{-3}(I_i - f)(u - C_i^1, v - C_i^2)\Big)\cdot\nabla f - \alpha\Delta Z = 0. \tag{233}$$

Rigorously speaking, it would be necessary to include a visibility function in the previous expression to account for the fields of views of the N_c cameras.

Observe that (233) only depends on the image values, f and $\nabla f = (f_u, f_v)^\top$, but not on the image gradients in ∇I_i, as already announced, making the resulting gradient descent flow more robust to image noise.

E.3.1 Boundary condition for the PDE in the height of the surface

The PDE (233) comes with natural boundary condition (34) because the surface is not closed. The geometric regularizer (28) yields the directional derivative of Z along $\boldsymbol{\nu}$, the unit normal to the integration domain U in the parameter space, as in (189).

$$(L_{\text{geom}})_{Z_u}\nu^u + (L_{\text{geom}})_{Z_v}\nu^v = Z_u\nu^u + Z_v\nu^v = \frac{\partial Z}{\partial \boldsymbol{\nu}}. \tag{234}$$

Now, formulas from section E.3 are used to compute such boundary condition for the data fidelity term. To simplify the calculations using the chain rule, substitute equations (199) and (200) in the left hand side of (34).

$$L_{Z_u}\nu^u + L_{Z_v}\nu^v = (\nu^u L_{\mathbf{X}_u} + \nu^v L_{\mathbf{X}_v}) \cdot \mathbf{e}_3. \tag{235}$$

In particular, using (215) and (217) for the data fidelity term, it follows that

$$\begin{aligned}
&(L_i)_{Z_u}\nu^u + (L_i)_{Z_v}\nu^v \\
&= \big(\nu^u (L_i)_{\mathbf{X}_u} + \nu^v (L_i)_{\mathbf{X}_v}\big) \cdot \mathbf{e}_3 \\
&= -\hat{\phi}_i |\mathtt{M}^i| \tilde{Z}_i^{-3} \Big(\nu^u \big(\mathbf{X}_v \times (\mathbf{X} - \mathbf{C}_i)\big) + \nu^v \big((\mathbf{X} - \mathbf{C}_i) \times \mathbf{X}_u\big)\Big) \cdot \mathbf{e}_3 \\
&= -\hat{\phi}_i |\mathtt{M}^i| \tilde{Z}_i^{-3} \big(\nu^u \det(\mathbf{e}_3, \mathbf{X}_v, \mathbf{X} - \mathbf{C}_i) + \nu^v \det(\mathbf{e}_3, \mathbf{X} - \mathbf{C}_i, \mathbf{X}_u)\big),
\end{aligned} \tag{236}$$

and, in a coordinate system adapted to the graph of the surface, it simplifies to

$$(L_i)_{Z_u}\nu^u + (L_i)_{Z_v}\nu^v = \phi_i |\mathtt{M}^i| \tilde{Z}_i^{-3} \big(\nu^u (u - C_i^1) + \nu^v (v - C_i^2)\big). \tag{237}$$

Combining (234) and (237) in (34) implies the boundary condition (38):

$$\sum_{i=1}^{N_c} \phi_i |\mathtt{M}^i| \tilde{Z}_i^{-3} \big((u - C_i^1)\nu^u + (v - C_i^2)\nu^v\big) + \alpha \frac{\partial Z}{\partial \boldsymbol{\nu}} = 0. \tag{238}$$

E.3.2 Alternative derivation of the variation with respect to the surface shape

Substituting (230) in (148) implies that the (unconstrained) gradient of the data-fidelity energy of the i-th image with respect to the surface shape is (ignoring here the boundary term that arises because the surface is not closed):

$$\nabla_S E_i = |\mathtt{M}^i| \tilde{Z}_i^{-3} (\hat{I}_i - \hat{f}) \big((\mathbf{X} - \mathbf{C}_i) \cdot \nabla \hat{f}\big)\, \mathbf{N}. \tag{239}$$

Another path to arrive at this result is to notice that E_i in (23) is written in a convenient way (145) to apply the theorem for flows in divergence form (157). Letting $\Phi(\mathbf{X}, \mathbf{N}) = \mathbf{G}(\mathbf{X}) \cdot \mathbf{N}$, with

$$\mathbf{G}(\mathbf{X}) = -|\mathtt{M}^i| \, \hat{\phi}_i(\mathbf{X}) \, \tilde{Z}_i^{-3}(\mathbf{X}) \, (\mathbf{X} - \mathbf{C}_i)$$

and, since $\hat{f}$ within $\hat{\phi}_i$ is fixed while computing the variation with respect to $\mathbf{X}$, the gradient $\nabla_S E_i$ is in divergence form (157):

$$\nabla_S E_i = (\nabla \cdot \mathbf{G}) \, \mathbf{N}, \tag{240}$$

with

$$\begin{aligned}
\nabla \cdot \mathbf{G} &= \nabla \cdot \big(-|\mathtt{M}^i| \, \hat{\phi}_i(\mathbf{X}) \, \tilde{Z}_i^{-3}(\mathbf{X}) \, (\mathbf{X} - \mathbf{C}_i)\big) \\
&= -|\mathtt{M}^i| \big(\hat{\phi}_i \tilde{Z}_i^{-3} \, \nabla \cdot (\mathbf{X} - \mathbf{C}_i) + (\mathbf{X} - \mathbf{C}_i) \cdot \nabla(\hat{\phi}_i \tilde{Z}_i^{-3})\big) \\
&= -|\mathtt{M}^i| \big(3\hat{\phi}_i \tilde{Z}_i^{-3} + (\mathbf{X} - \mathbf{C}_i) \cdot (\tilde{Z}_i^{-3} \nabla \hat{\phi}_i - 3\hat{\phi}_i \tilde{Z}_i^{-4} \mathbf{n}_3^i)\big) \\
&= -|\mathtt{M}^i| \Big(3\hat{\phi}_i \tilde{Z}_i^{-3} \big(1 - \tilde{Z}_i^{-1} (\mathbf{X} - \mathbf{C}_i) \cdot \mathbf{n}_3^i\big) + \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot \nabla \hat{\phi}_i \Big) \\
&= -|\mathtt{M}^i| \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot \nabla \hat{\phi}_i \\
&= -|\mathtt{M}^i| \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot \dot{\hat{\phi}}_i \, (\nabla \hat{I}_i - \nabla \hat{f}) \\
&= |\mathtt{M}^i| \tilde{Z}_i^{-3} \dot{\hat{\phi}}_i \, (\mathbf{X} - \mathbf{C}_i) \cdot \nabla \hat{f}.
\end{aligned} \tag{241}$$

The depth formula (180) and the constancy of the intensity along projection rays (214) were applied in the derivation to simplify the expression of the divergence. Observe that the term in ϕ_i cancels out and only the term in its derivative survives, as seen in the longer proof in appendix E.3 for the graph representation of the surface. In particular, if the photometric criterion is quadratic (15), formula (240) with (241) simplifies to (239). Finally, applying the results in appendix C for a multi-camera setup one may compute the gradient of E_{data} (corresponding to the interior of U) with respect to the surface height and arrive at (233).

There is yet another method to compute variation of the energy with respect to the surface shape, as mentioned in appendix B.2: the time derivative of the energy resulting from augmenting the unknowns with a fictitious time parameter. This approach is not shown in this thesis to avoid overwhelming the reader with repeated equations.

APPENDIX F

VON NEUMANN STABILITY ANALYSIS

Consider the stability in ℓ^2 of a time-stepping linear PDE of the form:

$$f_t = \beta \Delta f - cf + K. \tag{242}$$

For stability (propagation of errors with respect to perturbations of the initial solution) it is only necessary to consider the homogeneous equation ($K = 0$). Next, discretize (242) using forward difference in time and central differences in space (FTCS), yielding an explicit updating scheme. Under the usual discretization notation: $f^n_{i,j} = f(i\Delta x, j\Delta y, n\Delta t)$, similarly for $c(x, y)$, and assuming that the discretization grid is uniform in space ($\Delta x = \Delta y$), it follows that

$$\frac{f^{n+1}_{i,j} - f^n_{i,j}}{\Delta t} \approx \frac{\beta}{\Delta x^2}(f^n_{i+1,j} + f^n_{i-1,j} + f^n_{i,j+1} + f^n_{i,j-1} - 4f^n_{i,j}) - c_{i,j} f^n_{i,j},$$

that is, the explicit updating scheme:

$$f^{n+1}_{i,j} \approx f^n_{i,j} + \frac{\beta \Delta t}{\Delta x^2}(f^n_{i+1,j} + f^n_{i-1,j} + f^n_{i,j+1} + f^n_{i,j-1} - 4f^n_{i,j}) - \Delta t\, c_{i,j} f^n_{i,j}.$$

Since instabilities normally arise as local events in space and then propagate to the rest of the nodes in the discretized grid, to proceed with the analysis one may assume that $c_{i,j}$ is approximately constant $\bar{c}$ in the neighborhood under study. Now, taking spatial, discrete Fourier transforms on both sides of the equation[1] (hence, assuming spatial periodicity of the discrete function f_{ij}), it follows that

$$F^{n+1}(\omega_1, \omega_2) \approx F^n(\omega_1, \omega_2)\big(1 + \frac{\beta \Delta t}{\Delta x^2}(2\cos\omega_1 + 2\cos\omega_2) - \frac{4\beta \Delta t}{\Delta x^2} - \Delta t \bar{c}\big),$$

where the Fourier symbol of the PDE, i.e., the amplification factor that measures the growth or decay of a Fourier mode of the error during an iteration, is

$$\rho(\omega_1, \omega_2) = 1 + \frac{2\beta \Delta t}{\Delta x^2}(\cos\omega_1 + \cos\omega_2) - \Delta t(\frac{4\beta}{\Delta x^2} + \bar{c}).$$

[1] For example, use the symmetric definition of the F.T.: $X(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} x_k e^{-\imath \omega k}$, where $\imath = \sqrt{-1}$. The inverse F.T. is $x_k = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} X(\omega) e^{\imath \omega k} \mathrm{d}\omega$.

The condition for ℓ^2 stability of the PDE is that $|\rho(\omega_1, \omega_2)| \leq 1$ for all frequencies, that is, the Fourier must not amplify the errors. For a real symbol, this implies $-1 \leq \rho \leq 1$. Taking into account that the maximum and minimum values that $(\cos\omega_1 + \cos\omega_2)$ may achieve are 2 and -2, respectively, it follows that

$$\begin{aligned}
\rho \leq 1 \quad &\Rightarrow \quad \frac{2\beta\Delta t}{\Delta x^2}(\cos\omega_1 + \cos\omega_2) - \Delta t(\frac{4\beta}{\Delta x^2} + \bar{c}) \leq 0 \\
&\Rightarrow \quad \frac{4\beta\Delta t}{\Delta x^2} - \Delta t(\frac{4\beta}{\Delta x^2} + \bar{c}) \leq 0 \\
&\Leftrightarrow \quad -c\bar{\Delta} t \leq 0 \\
&\Leftrightarrow \quad \bar{c} \geq 0,
\end{aligned}$$

$$\begin{aligned}
\rho \geq -1 \quad &\Rightarrow \quad \frac{2\beta\Delta t}{\Delta x^2}(\cos\omega_1 + \cos\omega_2) - \Delta t(\frac{4\beta}{\Delta x^2} + \bar{c}) \geq -2 \\
&\Rightarrow \quad \frac{-4\beta\Delta t}{\Delta x^2} - \Delta t(\frac{4\beta}{\Delta x^2} + \bar{c}) \geq -2 \\
&\Leftrightarrow \quad \Delta t(\frac{4\beta}{\Delta x^2} + \frac{\bar{c}}{2}) \leq 1 \\
&\Leftrightarrow \quad \Delta t \leq (\frac{4\beta}{\Delta x^2} + \frac{\bar{c}}{2})^{-1}.
\end{aligned}$$

Interpreting both previous conditions in the entire domain (not just as a local analysis), they imply that for ℓ^2 stability in the worst case scenario it is required that

$$c_{i,j} \geq 0 \ \ \forall i,j \qquad \text{and} \qquad \Delta t \leq \big(\frac{4\beta}{\Delta x^2} + \frac{c_{i,j}}{2}\big)^{-1} \ \ \forall i,j,$$

from which (52) follows.

Among the many assumptions, there is yet another one: Fourier analysis assumes periodic boundary conditions, too, which is not always the case in a real life situation. Nevertheless, there is no other easy method to derive a stability condition.

F.1 Stability condition for the radiance PDE

The previous result can be applied to state the ℓ^2 stability condition for the gradient descent PDE in the radiance:

$$f_t = \beta\Delta f - (\sum_{k=1}^{N_c} \mathtt{J}_{i,j}^k) f + \sum_{k=1}^{N_c} I_{i,j}^k \mathtt{J}_{i,j}^k.$$

Identifying coefficients (53), the stability conditions (52) are implied.

F.2 Stability condition for the height PDE with Laplacian diffusive term

The gradient descent PDE for the height function is (43). Considering a general photometric error function ϕ, the PDE has a non-linear term of the form (see (231) and (241))

$$g(Z) = \sum_{k=1}^{N_c} |\mathtt{M}^k| \tilde{Z}_k^{-3} \dot{\hat{\phi}}_k (\mathbf{X} - \mathbf{C}_k) \cdot \nabla \hat{f}, \tag{243}$$

where $\dot{\hat{\phi}}_k \equiv \dot{\hat{\phi}}_k(\hat{I}_k - \hat{f})$ is the derivative of the extension of the photometric criterion, as in (195). In a coordinate system adapted to the graph, $\mathbf{X} = (u, v, Z(u,v))^\top$ and the non-linear term simplifies to (41):

$$g(Z) = \sum_{k=1}^{N_c} |\mathtt{M}^k| \tilde{Z}_k^{-3} \dot{\phi}_k (u - C_k^1, v - C_k^2) \cdot \nabla f, \tag{244}$$

with $\dot{\phi}_k \equiv \dot{\phi}_k(I_k - f)$. Using a Taylor series expansion around the current height $\bar{Z}$, we may linearize the non-linear term as in (49), $g(Z) \approx g(\bar{Z}) + \dot{g}(\bar{Z})(Z - \bar{Z})$, and consequently the gradient descent PDE is also in linearized form:

$$Z_t \approx \alpha \Delta Z - g(\bar{Z}) - \dot{g}(\bar{Z})(Z - \bar{Z}).$$

We may now apply the previously derived stability result for linear PDEs (242). Identifying terms between both linear PDEs, we have

$$\beta = \alpha, \quad c = \dot{g}(\bar{Z}) \quad \text{and} \quad K = -g(\bar{Z}) + \dot{g}(\bar{Z})\bar{Z},$$

yielding the stability condition (54).

F.3 Derivative of the non-linear term in the height PDE

An expression for $\dot{g}(\bar{Z})$ is needed to compute the maximum time step in (54). Calculations are carried out in both coordinate systems: adapted to the graph (henceforth, graph frame) and not adapted, i.e., a general world or inertial frame, stemming from the expressions of the non-linear term (243) and (244).

Let $\mathbf{X}' = \big(X'(u,v), Y'(u,v), Z'(u,v)\big)^\top$ be the inertial coordinates of a surface point and let $\mathbf{X} = \big(u, v, Z(u,v)\big)^\top$ be its coordinates with respect to a frame adapted to the graph

(e.g. graph coordinates). Both representations are related by a Euclidean transformation of space (i.e., a rigid body motion): $\mathbf{X}' = \mathtt{Q}\mathbf{X} + \mathbf{t}_0$, where $\mathtt{Q} = (\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$ is a rotation matrix and $\mathbf{t}_0$ is a translation vector. Geometrically, $\mathbf{t}_0$ are the inertial coordinates of the origin of the graph frame and the columns of $\mathtt{Q}$ are the inertial coordinates of the basis vectors of the graph frame. In particular, $\mathbf{q}_3$ represents the direction of variation of the height of the graph, z: $\mathbf{X}'_z = \mathtt{Q}\mathbf{X}_z = \mathtt{Q}\mathbf{e}_3 = \mathbf{q}_3$. The transformation rule for free vectors between both coordinate frames is $\mathbf{w}' = \mathtt{Q}\mathbf{w}$.

Since $\tilde{Z}_k$ is the depth of a surface point with respect to the k-th camera and, intuitively, this quantity is the same regardless of the coordinate frame, we drop the prime notation for this term. Recall that the inner product is invariant to orthogonal transformations, therefore if $\mathbf{n}_3'^k = \mathtt{Q}\mathbf{n}_3^k$: $\tilde{Z}'_k = (\mathbf{X}' - \mathbf{C}'_k) \cdot \mathbf{n}_3'^k = (\mathbf{X} - \mathbf{C}_k) \cdot \mathbf{n}_3^k = \tilde{Z}_k$.

The extension of the radiance function is defined to be constant along the direction of variation of the height (see section E.2), i.e., in graph coordinates: $\hat{f}(\mathbf{X}) \equiv \hat{f}(u, v, Z) = f(u, v)$. In arbitrary coordinates (prime notation), for consistency, the extension $\hat{f}' : \mathbb{R}^3 \to \mathbb{R}$ is defined by: $\hat{f}'(\mathbf{X}') = \hat{f}(\mathbf{X})$. A similar definition applies to the extension of the image intensity along projection rays: $\hat{I}'_k(\mathbf{X}') = \hat{I}_k(\mathbf{X})$. This implies, through the chain rule, that the derivative

$$D\hat{f}(\mathbf{X}) = D\hat{f}'(\mathbf{X}') \, \frac{\partial \mathbf{X}'}{\partial \mathbf{X}} = D\hat{f}'(\mathbf{X}') \, \mathtt{Q}$$

or, in column notation,$\nabla \hat{f}(\mathbf{X}) = \mathtt{Q}^\top \nabla \hat{f}'(\mathbf{X}')$. The relationship between second order derivatives $D^2 \equiv \nabla^2$ in both frames is:

$$D^2 \hat{f}(\mathbf{X}) = D\big(\mathtt{Q}^\top \nabla \hat{f}'(\mathbf{X}')\big) = \mathtt{Q}^\top D^2 \hat{f}'(\mathbf{X}') \, \mathtt{Q},$$

and by simple calculations: $D\hat{f}(\mathbf{X}) = \big(f_u, f_v, 0\big) = (\nabla f^\top, 0)$,

$$D^2 \hat{f}(\mathbf{X}) = \begin{pmatrix} f_{uu} & f_{uv} & 0 \\ f_{uv} & f_{vv} & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} D^2 f & 0 \\ 0 & 0 \end{pmatrix},$$

where f_u, f_v, f_{uu}, etc. are functions of (u, v). Whenever it is clear from the context, we will simplify the notation so that previous relations become: $\nabla \hat{f} = \mathtt{Q}^\top \nabla \hat{f}'$ and $D^2 \hat{f} = \mathtt{Q}^\top D^2 \hat{f}' \, \mathtt{Q}$.

Differentiating (243) with respect to the height Z, yields

$$\begin{aligned}\dot{g}(Z) &= \sum_{k=1}^{N_c} |\mathtt{M}^k| \Big(-3\tilde{Z}_k^{-4} (\mathbf{X}'_z \cdot \mathbf{n}_3'^k)\, \dot{\hat{\phi}}_k(\hat{I}'_k - \hat{f}')\, (\mathbf{X}' - \mathbf{C}'_k) \cdot \nabla \hat{f}' \\ &\quad + \tilde{Z}_k^{-3} \big(\ddot{\hat{\phi}}_k (\hat{I}'_k - \hat{f}')\, (\nabla \hat{I}'_k - \nabla \hat{f}') \cdot \mathbf{X}'_z \big) (\mathbf{X}' - \mathbf{C}'_k) \cdot \nabla \hat{f}' \\ &\quad + \tilde{Z}_k^{-3}\, \dot{\hat{\phi}}_k (\hat{I}'_k - \hat{f}')\, \big((\nabla \hat{f}')^\top \mathbf{X}'_z + (\mathbf{X}' - \mathbf{C}'_k)^\top (\nabla^2 \hat{f}')\, \mathbf{X}'_z \big) \Big).\end{aligned}$$

Recalling the gauge freedom in the definition of the extension $\hat{f}$ and the properties of the inner product, some simplifications occur. The last term in the previous equation vanishes and the second term simplifies:

$$\begin{aligned}\mathbf{X}'_z \cdot \mathbf{n}_3'^k &= \mathbf{q}_3 \cdot \mathbf{n}_3'^k \\ (\nabla \hat{f}')^\top \mathbf{X}'_z &= (\nabla \hat{f})^\top \mathtt{Q}^\top \mathtt{Q} \mathbf{e}_3 = (\nabla \hat{f})^\top \mathbf{e}_3 = (\nabla f^\top, 0) \cdot (0,0,1) = 0 \\ (\nabla \hat{I}'_k - \nabla \hat{f}') \cdot \mathbf{X}'_z &= \nabla \hat{I}'_k \cdot \mathbf{q}_3 - \nabla \hat{f}' \cdot \mathbf{q}_3 = \nabla \hat{I}'_k \cdot \mathbf{q}_3 - \nabla \hat{f} \cdot \mathbf{e}_3 = \nabla \hat{I}'_k \cdot \mathbf{q}_3 \\ (\mathbf{X}' - \mathbf{C}'_k)^\top (\nabla^2 \hat{f}')\, \mathbf{X}'_z &= (\mathbf{X} - \mathbf{C}_k)^\top \mathtt{Q}^\top (\nabla^2 \hat{f}') \mathtt{Q}\, \mathbf{e}_3 = (\mathbf{X} - \mathbf{C}_k)^\top (\nabla^2 \hat{f})\, \mathbf{e}_3 = 0.\end{aligned}$$

Therefore, in the general frame the desired expression is:

$$\dot{g}(Z) = \sum_{k=1}^{N_c} |\mathtt{M}^k| \tilde{Z}_k^{-3} \Big(\big(-3\tilde{Z}_k^{-1}\, \dot{\hat{\phi}}_k (\hat{I}'_k - \hat{f}')\, \mathbf{n}_3'^k + \ddot{\hat{\phi}}_k (\hat{I}'_k - \hat{f}')\, \nabla \hat{I}'_k \big) \cdot \mathbf{q}_3 \Big) (\mathbf{X}' - \mathbf{C}'_k) \cdot \nabla \hat{f}' \quad (245)$$

In the graph frame, we may use the fact that $\mathbf{n}_3^k \cdot \mathbf{e}_3 = \mathtt{R}_{33}^k$ and substitute the definition of the extensions $\hat{f}$, $\hat{I}_k$ in

$$(\mathbf{X}' - \mathbf{C}'_k) \cdot \nabla \hat{f}' = (\mathbf{X} - \mathbf{C}_k) \cdot \nabla \hat{f} = (u - C_k^1, v - C_k^2, Z - C_k^3) \cdot (\nabla f^\top, 0) = (u - C_k^1, v - C_k^2) \cdot \nabla f$$

to further simplify the expression for derivative of the non-linear term:

$$\dot{g}(Z) = \sum_{k=1}^{N_c} |\mathtt{M}^k| \tilde{Z}_k^{-3} \Big(\big(-3\tilde{Z}_k^{-1}\, \dot{\phi}_k (I_k - f)\, \mathtt{R}_{33}^k + \ddot{\phi}_k (I_k - f)\, \nabla \hat{I}_k \cdot \mathbf{e}_3 \big) (u - C_k^1, v - C_k^2) \cdot \nabla f. \quad (246)$$

In particular, for the quadratic photometric error function in (15), (43):

$$\dot{g}(Z) = \sum_{k=1}^{N_c} |\mathtt{M}^k| \tilde{Z}_k^{-3} \Big(\big(-3\tilde{Z}_k^{-1} (I_k - f) \mathtt{R}_{33}^k + \nabla \hat{I}_k \cdot \mathbf{e}_3 \big) (u - C_k^1, v - C_k^2) \cdot \nabla f. \quad (247)$$

F.4 Stability condition for the height PDE with Mean curvature diffusive term

To derive a stability condition for the PDE

$$\begin{aligned} Z_t &= \big(\alpha 2H - g(Z)\big)/\sqrt{1+Z_u^2+Z_v^2} \\ &= \alpha\frac{(1+Z_v^2)Z_{uu} - 2Z_uZ_vZ_{uv} + (1+Z_u^2)Z_{vv}}{(1+Z_u^2+Z_v^2)^2} - \frac{g(Z)}{\sqrt{1+Z_u^2+Z_v^2}}, \end{aligned}$$

three assumptions are exploited: (i) the second order derivatives of z vary more quickly than the first order derivatives, so that the latter can be locally approximated by constants, (ii) instabilities normally arise as local events and propagate to the rest of the grid and (iii) the non-linear term can be well approximated by its first order Taylor expansion.

Next, discretize the linearized, homogeneous PDE with the new diffusive term (mean curvature) from the minimal area regularizer and repeat the steps of the ℓ^2 stability analysis in the Fourier domain (von Neumann analysis). Subsuming the factor $\sqrt{1+Z_u^2+Z_v^2}$ in H and $g(Z)$, and noting the new terms with a prime notation, one will obtain the explicit scheme:

$$Z_{i,j}^{n+1} \approx Z_{i,j}^n + \Delta t(\alpha 2H'_{i,j} - \dot{g}'(\bar{Z})Z_{i,j}^n), \tag{248}$$

where the mean curvature term is locally approximated and discretized by

$$\begin{aligned} 2H'_{i,j} &= (AZ_{uu} + BZ_{vv} + CZ_{uv})_{i,j} \\ &\approx \frac{1}{\Delta x^2}\big(A_{i,j}(Z_{i+1,j} - 2Z_{i,j} + Z_{i-1,j}) + B_{i,j}(Z_{i,j+1} - 2Z_{i,j} + Z_{i,j-1}) \\ &\qquad + \frac{C_{i,j}}{4}(Z_{i+1,j+1} + Z_{i-1,j-1} - Z_{i+1,j-1} - Z_{i-1,j+1})\big), \end{aligned}$$

with

$$A = \frac{1+Z_v^2}{(1+Z_u^2+Z_v^2)^2}, \quad B = \frac{1+Z_u^2}{(1+Z_u^2+Z_v^2)^2}, \quad C = \frac{-2Z_uZ_v}{(1+Z_u^2+Z_v^2)^2},$$

and using central differences for discretizing the second order derivatives of z:

$$\begin{aligned} Z_{uu} &\approx \frac{1}{\Delta x^2}(Z_{i+1,j} - 2Z_{i,j} + Z_{i-1,j}) \\ Z_{vv} &\approx \frac{1}{\Delta x^2}(Z_{i,j+1} - 2Z_{i,j} + Z_{i,j-1}) \\ Z_{uv} &\approx \frac{1}{4\Delta x^2}(Z_{i+1,j+1} + Z_{i-1,j-1} - Z_{i+1,j-1} - Z_{i-1,j+1}). \end{aligned}$$

Taking the Fourier transform on (248), we get $Z^{n+1}(\omega_1,\omega_2) = Z^n(\omega_1,\omega_2)\rho^n(\omega_1,\omega_2)$, where the Fourier symbol $\rho^n(\omega_1,\omega_2)$ is

$$1+\Delta t\Big(-\big(\dot g'(\bar Z)\big)^m+\frac{2\alpha}{\Delta x^2}\big(A^m(\cos\omega_1-1)+B^m(\cos\omega_2-1)+\tilde C^m(\cos(\omega_1+\omega_2)-\cos(\omega_1-\omega_2))\big)\Big),$$

with $\tilde C = C/4$. The magnitude constraint on the amplification factor ($|\rho| \le 1$) yields the stability conditions $-1 \le \rho^n \le 1$. The second inequality, $\max\rho^n - 1 \le 0$ is satisfied if

$$0 \le \frac{\alpha}{\Delta x^2} \le \min \dot g(\bar Z^n_{i,j}) \tag{249}$$

Proof. Impose the condition $\max\rho^n - 1 \le 0$ by requiring an upper bound on $(\max\rho^n - 1)$ to be ≤ 0. Since $\max(a+b) \le \max(a) + \max(b)$ and $\max(-a) = -\min(a)$,

$$\begin{aligned}
\max\rho^n - 1 &\le \\
-\min\big(\dot g'(\bar Z)\big)^n + \frac{2\alpha}{\Delta x^2}(A^n\cdot 0 + B^n\cdot 0 + \max 2|\tilde C^n|) &= \\
-\min\big(\dot g'(\bar Z)\big)^n + \frac{\alpha}{\Delta x^2}\max|C^n| &= \\
-\min\Big(\frac{\dot g(\bar Z)}{\sqrt{1+Z_u^2+Z_v^2}}\Big)^n + \frac{\alpha}{\Delta x^2}\frac{2}{(1+Z_u^2+Z_v^2)} &\le 0,
\end{aligned}$$

that is,

$$\frac{\alpha}{\Delta x^2}\frac{2}{\sqrt{1+Z_u^2+Z_v^2}} \le \min\big(\dot g(\bar Z)\big)^n.$$

Now, undoing the local constancy assumption and considering the max/min over the entire numerical grid at time the current time $n\Delta t$, one gets (249):

$$\frac{2\alpha}{\Delta x^2}\max\frac{1}{\sqrt{1+(Z_u^2+Z_v^2)^n_{i,j}}} \le \frac{2\alpha}{\Delta x^2} \le \min\dot g(\bar Z^n_{i,j}). \qquad \square$$

The first inequality, $-1 \le \min\rho^n$ is satisfied if

$$\Delta t \le \Big(\frac{1}{2}\max|\dot g(\bar Z^n_{i,j})| + \frac{5\alpha}{\Delta x^2}\Big)^{-1}. \tag{250}$$

Proof. Impose the condition $-1 \le \min\rho^n$ by requiring a lower bound on $(\min\rho^n - 1)$ to be ≥ -2. Since $\min(a+b) \ge \min(a) + \min(b)$ and $\min(-a) = -\max(a)$,

$$\begin{aligned}
(\min\rho^n - 1)/\Delta t &\ge \\
-\max\big(\dot g'(\bar Z)\big)^n + \frac{2\alpha}{\Delta x^2}(-2\max A^n - 2\max B^n - 2\max|\tilde C^n|) &= \\
-\max(\dot g'(\bar Z))^n - \frac{4\alpha}{\Delta x^2}(\max A^n + \max B^n + \max|\tilde C^n|) &= \\
-\max\Big(\frac{\dot g(\bar Z)}{\sqrt{1+Z_u^2+Z_v^2}}\Big)^n - \frac{4\alpha}{\Delta x^2}\frac{5}{2(1+Z_u^2+Z_v^2)} &\ge \frac{-2}{\Delta t},
\end{aligned}$$

that is,

$$\frac{1}{2}\max\left(\frac{\dot{g}(\bar{Z})}{\sqrt{1+Z_u^2+Z_v^2}}\right)^n + \frac{\alpha}{\Delta x^2}\frac{5}{(1+Z_u^2+Z_v^2)} \leq \frac{1}{\Delta t}.$$

Undoing the local constancy assumption by considering the max/min over the entire numerical grid at time the current time $n\Delta t$, yields (250). In the derivation, we used the loose bounds $0 \leq \{A^n, B^n, |C^n|/2\} \leq (1+Z_u^2+Z_v^2)^{-1}$ so that

$$A + B + |\tilde{C}| \leq (5/2)(1+Z_u^2+Z_v^2)^{-1}. \qquad \square$$

APPENDIX G

MEAN PLANE OF A SET OF SPACE POINTS

As already mentioned in chapter 4, one challenging problem in our paired oceanic application is converting camera coordinates to true world coordinates solely from the images because the latter lack information about known 3-D landmarks. The above problem consists of two decoupled subproblems: (i) compute the gravity plane in the same frame as the cameras are given (usually, camera coordinates) and (ii) reorient the gravity plane to place it in its rightful position, $Z = 0$, in the world frame. It is clear that if the cameras are given in the world frame, then there is no need to estimate the gravity plane since it already coincides with the plane $Z = 0$. Often, cameras' extrinsic parameters (orientation and translation) are only specified in a relative manner with respect to a reference camera, therefore the previous ideal situation of coincident frames never occurs.

Assuming the ocean surface can be represented by a collection of scattered 3-D points (i.e., a point cloud), different methods to estimate the mean plane through this set of points have been studied to determine the gravity plane. Once this plane has been obtained, a Euclidean transformation (i.e., a rigid body motion) can be specified to convert camera coordinates to world coordinates. A procedure will be given to determine such transformation. In the following, assume there are no outliers in the set of input 3-D points, otherwise, the methods presented can be used in conjunction with robust methods such as RANdom SAmple Consensus (RANSAC) [24] or Least MEDian of Squares (LMEDS) [113] to deal with this type of noise.

The mean plane through the scene is estimated using finite-dimensional optimization methods since a plane can be parameterized by its 4 coefficients (homogeneous coordinates)

$$\mathbf{v} = (a, b, c, d)^\top \tag{251}$$

so that a 3-D point with homogeneous coordinates $\mathbf{X} = (X, Y, Z, 1)^\top$ lies in the plane $\mathbf{v}$ if

Given $n \geq 3$ points in space with coordinates $\{(X_i, Y_i, Z_i)\}_{i=1}^n$,

- Build the $n \times 4$ matrix A whose i-th row is $(X_i, Y_i, Z_i, 1)$.
- Compute the Singular Value Decomposition (SVD) of A, i.e., $A = UDV^\top$.
- The plane $\mathbf{v}$ that minimizes the algebraic distance to the points, $\|A\mathbf{v}\|/\|\mathbf{v}\|$, is the right singular vector of A associated to the smallest singular value, i.e., the last column of V assuming that the singular values in D are sorted in descending order.

Figure 51: Algorithm to fit a plane to a set of points by minimizing the algebraic distance from the points to the plane.

and only if $\mathbf{X}^\top \mathbf{v} = aX + bY + cZ + d = 0$. Several cost functions are discussed.

Minimization of the algebraic distance. This method is summarized in Figure 51 and it estimates the plane $\mathbf{v}$ by minimizing an *algebraic* cost function. Based on the equation $\mathbf{X}^\top \mathbf{v} = 0$, the coordinates of the plane that passes through $n \geq 3$ given points $\{\mathbf{X}_i\}_{i=1}^n$ can be found by solving the linear system of equations $A\mathbf{v} = 0$, where A is an $n \times 4$ matrix whose i-th row is $\mathbf{X}_i^\top$. However, if the coordinates of the points are noisy, each point will provide an equation of the form $\mathbf{X}_i^\top \mathbf{v} = \epsilon_i$, which no longer leads to a homogeneous system of equations, but to a system $A\mathbf{v} = \mathbf{r}$, with $\mathbf{r} = (\epsilon_1, \ldots, \epsilon_n)^\top$, that might not have an exact solution. Using least squares, one finds $\mathbf{v}$ by minimizing the so called algebraic distance $\|A\mathbf{v}\|$. Since $\mathbf{v}$ is only determined up to a scale factor, the additional constraint $\|\mathbf{v}\| = 1$ is enforced. The estimation problem $\min \|A\mathbf{v}\|^2$ subject to $\|\mathbf{v}\| = 1$ is classical. Its solution $\mathbf{v}$ is the eigenvector of $A^\top A$ associated to the smallest eigenvalue of $A^\top A$, or equivalently, the right singular vector of A associated to the smallest singular value of A.

Minimization of the perpendicular Euclidean distance The method summarized in Figure 52 estimates the plane v by minimizing a *geometric* cost function: the sum of squares of perpendicular distances from the data points to the plane.

The signed perpendicular (orthogonal) distance from a point $\mathbf{X}$ to a plane $\mathbf{v}$ is

$$sd(\mathbf{X}, \mathbf{v}) = \frac{\mathbf{X}^\top \mathbf{v}}{\sqrt{a^2 + b^2 + c^2}} = \frac{aX + bY + cZ + d}{\sqrt{a^2 + b^2 + c^2}}$$

and the true distance is its absolute value, $d(\mathbf{X}, \mathbf{v}) = |sd(\mathbf{X}, \mathbf{v})|$. To emphasize that it is the

Given $n \geq 3$ points in space with coordinates $\{(X_i, Y_i, Z_i)\}_{i=1}^n$,

- Build the $n \times 4$ matrix A whose i-th row is $\mathbf{X}_i^\top = (X_i, Y_i, Z_i, 1)$.
- Compute the solution to the Generalized Eigenvalue Problem $A^\top A\mathbf{v} = \lambda M\mathbf{v}$, where $M = \text{diag}(1, 1, 1, 0)$.
- The plane $\mathbf{v} = (a, b, c, d)^\top$ that minimizes the sum of square perpendicular (Euclidean) distances from the points to the plane, $\sum_{i=1}^n d^2(\mathbf{X}_i, \mathbf{v})$, is the eigenvector associated to the smallest (positive) finite eigenvalue.

Figure 52: Algorithm to fit a plane to a set of points by minimizing the sum of square perpendicular distances from the points to the plane.

(Euclidean) perpendicular distance, the notation $d_\perp(\mathbf{X}, \mathbf{v})$ is also used. A sensible criterion to find a plane $\mathbf{v}$ through a set of $\{\mathbf{X}_i\}_{i=1}^n$ points, is to minimize the geometric error

$$\sum_{i=1}^n d^2(\mathbf{X}_i, \mathbf{v}). \tag{252}$$

Since the coordinates of the plane $\mathbf{v}$ are determined up to a scale factor, the additional constraint of unit normal $\|\mathbf{n}\| = 1$, where $\mathbf{n} = (a, b, c)^\top$, eliminates such degree of freedom. Therefore, the optimization problem is $\min_{\mathbf{v}} \sum_{i=1}^n (\mathbf{X}_i^\top \mathbf{v})^2$ subject to $a^2 + b^2 + c^2 = 1$. In matrix form, this becomes $\min_{\mathbf{v}} \mathbf{v}^\top A^\top A\mathbf{v}$ subject to $\mathbf{v}^\top M\mathbf{v} = 1$, where $M = \text{diag}(1, 1, 1, 0)$ and A is an $n \times 4$ matrix whose i-th row is $\mathbf{X}_i^\top$. This problem can be solved using the method of Lagrange multipliers: $\mathbf{v}$ is a solution of the generalized eigenvalue problem $A^\top A\mathbf{v} = \lambda M\mathbf{v}$. The cost associated to $\mathbf{v}$ is $\mathbf{v}^\top A^\top A\mathbf{v} = \lambda \mathbf{v}^\top M\mathbf{v} = \lambda$. Therefore, the solution plane $\mathbf{v}$ that minimizes the cost function is the eigenvector corresponding to the smallest (positive) finite eigenvalue λ. It can be proved that the plane estimated trough this method is a "mean" plane through the data, i.e.,

$$\sum_{i=1}^n sd(\mathbf{X}_i, \mathbf{v}) = 0. \tag{253}$$

Relationship between the algebraic and the geometric distances. The algebraic distance is related to the (geometric) orthogonal distance by

$$d_{\text{alg}}^2(\mathbf{X}, \mathbf{v}) = \frac{(\mathbf{X}^\top \mathbf{v})^2}{\mathbf{v}^\top \mathbf{v}} = d_\perp^2(\mathbf{X}, \mathbf{v}) \left(1 + \frac{d^2}{\|\mathbf{n}\|^2}\right)^{-1}.$$

Therefore, both distances are the same if the last coordinate of the estimated plane is $d = 0$, i.e., if the plane passes through the world origin $\mathbf{X}_0 = (0, 0, 0, 1)^\top$.

Given $n \geq 3$ points in space with coordinates $\{(X_i, Y_i, Z_i)\}_{i=1}^n$,

1. **Normalization of the points**. Apply a similarity transformation H_S to the points such that it moves the centroid of the points to the origin and makes the average distance of the transformed points to the origin equal to $\sqrt{3}$ (scaling for 3-D points). The rotation matrix is irrelevant: choose $R = I$ (identity matrix).

2. **Estimation of the plane**. Use the algorithm in Figure 51 on the transformed points to obtain the fitted plane $\mathbf{v}'$. Such algorithm can be further simplified because the last coefficient of the estimated vector $\mathbf{v}' = (a', b', c', d')^\top$ will be $d' = 0$.

3. **Denormalization of the plane**. Set $\mathbf{v} = H_S^\top \mathbf{v}'$. This plane minimizes the geometric error (252) with respect to the given data points.

Figure 53: A simple alternative to the algorithm in Figure 51 to fit a plane to a set of points.

Boosting methods by transformations. To improve the numerical conditioning of the matrices in the Least Squares problems considered, a similarity transformation (rotation, translation and isotropic scaling) can be applied to the data points before using any of the methods in Figures 51 and 52 to fit a plane. This transformation does not change the minimizer of the geometric error (252). However, it can have a dramatic positive effect on the minimizer of the algebraic error. In our benefit, choosing the right transformation can make the algebraic distance have the same minimizer as the geometric error.

A similarity transformation is represented by a 4×4 homogeneous matrix

$$H_S = \begin{pmatrix} \gamma R & \mathbf{t} \\ 0^\top & 1 \end{pmatrix},$$

where R is a rotation matrix, $\mathbf{t}$ is a vector and $\gamma \neq 0$ is a scalar. Points $\mathbf{X}$ and planes $\mathbf{v}$ are transformed according to the rules $\mathbf{X}' = (X', Y', Z', 1)^\top = H_S \mathbf{X}$ and $\mathbf{v}' = (a', b', c', d')^\top = H_S^{-\top} \mathbf{v}$ so that $\mathbf{v}^\top \mathbf{X} = \mathbf{v}'^\top \mathbf{X}'$. This transformation induces a scaling of the signed Euclidean distance: $sd(\mathbf{X}', \mathbf{v}') = |\gamma| \, sd(\mathbf{X}, \mathbf{v})$. As expected, the Euclidean distance is invariant to rotations and translations, but not to scaling of the coordinates.

Any similarity transformation that places the centroid of the transformed points at the origin will make the algebraic distance to be a scaled version of the geometric error. This suggests that a trivial pre-processing of the data can make a simple algorithm (Figure 51)

become very powerful (Figure 52). A simple yet powerful transformation H_S that provides good numerical conditioning is specified in Figure 53.

The methods in Figures 52 and 53 provide the best results in terms of (i) minimizing the perpendicular distance and (ii) mean plane (253) criterion. Algorithm in Figure 53 is the preferred method among the three discussed.

Plane reorientation. Assuming that the gravity plane $\mathbf{v} = (a, b, c, d)^\top$ is known, now the goal is to find the rigid body motion that moves $\mathbf{v}$ to its right location ($Z' = 0$) in the new coordinate system. From projective geometry, if points with homogeneous coordinates $\mathbf{X} = (X, Y, Z, 1)^\top$ transform as $\mathbf{X}' = (X', Y', Z', 1)^\top = H_E\mathbf{X}$, where

$$H_E = \begin{pmatrix} R & \mathbf{t} \\ 0^\top & 1 \end{pmatrix} \tag{254}$$

is the 4×4 matrix of the transformation between coordinate systems, then planes transform as $\mathbf{v}' = H_E^{-\top}\mathbf{v}$. Therefore, the goal is to specify the rotation matrix R and the translation vector $\mathbf{t}$ such that $\mathbf{v}' = (0, 0, 1, 0)^\top$ is the transformed of the estimated gravity plane $\mathbf{v}$. That is, if $\mathbf{v} = (\mathbf{n}^\top, d)^\top$ and $\mathbf{v}' = (\mathbf{n}'^\top, d')^\top$, then the above condition yields two equations:

$$\mathbf{n}' = (0, 0, 1)^\top = R\,\mathbf{n} \tag{255}$$

$$-\mathbf{t}^\top R\mathbf{n} + d = 0. \tag{256}$$

Observe that R is the mapping between unit normals to the planes (255). Therefore, the angle and axis of rotation are given by the cross product of $\mathbf{n}$ and $\mathbf{n}'$. Rodrigues' rotation formula can be used to compute R:

$$R = I + \sin\theta\,[w_u]_\times + (1 - \cos\theta)\,[w_u]_\times^2\,. \tag{257}$$

Given $\mathbf{n} = (a, b, c)^\top$ and $\mathbf{n}' = (0, 0, 1)^\top$, the rotation axis in (257) is $w_u = (\mathbf{n} \times \mathbf{n}')/(\|\mathbf{n} \times \mathbf{n}'\|) = \frac{1}{\sqrt{a^2+b^2}}(b, -a, 0)^\top$, and the rotation angle $0 \leq \theta \leq \pi$ is, from the geometrical interpretation of the dot product, $\theta = \arccos\frac{\mathbf{n}^\top\mathbf{n}'}{\|\mathbf{n}\|\cdot\|\mathbf{n}'\|} = \arccos(c)$. Substitute the previous

results and $q = (1-c)/(a^2+b^2)$ in (257) to obtain

$$R = \begin{pmatrix} 1-qa^2 & -qab & -a \\ -qab & 1-qb^2 & -b \\ a & b & c \end{pmatrix}$$

Observe that the rotation matrix only depends on the unit normal to the plane $\mathbf{v}$. It can be checked that $R\mathbf{n} = (0, 0, \|\mathbf{n}\|)^\top = (0, 0, 1)^\top$, as expected. Once the rotation R is known, the translation $\mathbf{t} = (t_1, t_2, t_3)^\top$ can be computed from (256). Since $R\mathbf{n} = (0, 0, 1)^\top$, only the third coordinate of $\mathbf{t}$ is needed: equation (256) becomes $t_3 = d$. The coordinates t_1 and t_2 are undetermined because they specify the location of an arbitrary world origin on the plane $Z' = 0$. We may choose $\mathbf{t} = (0, 0, d)^\top$.

The estimation methods provide two equally valid solutions: $\mathbf{v}$ and $-\mathbf{v}$, but only one of them is physically correct. This discrete ambiguity is resolved by choosing the plane $\mathbf{v}$ whose H_E transforms an extra control point as a priori known: for example, the optical center of a camera lies on a half-space with respect to the plane $\mathbf{v}$ and it must lie on the positive half space with respect to the plane $\mathbf{v}' \equiv Z = 0$. There is no need to convert the point since the half-space property is described by the signed distance to the plane and it is preserved by rigid body motions, thus the physically correct solution is the one with positive signed distance for the optical center of the camera.

Finally, one may apply the transformation H_E (254) to convert the surface points to world coordinates: $(X', Y', Z')^\top = R\,(X, Y, Z)^\top + \mathbf{t}$. If desired, an extra rotation around the Z' axis may be necessary to align the optical centers of the cameras along the X' axis.

REFERENCES

[1] Adalsteinsson, D. and Sethian, J. A., "A fast level set method for propagating interfaces," *J. Comput. Phys.*, vol. 118, no. 2, pp. 269–277, 1995.

[2] Agarwal, S., Furukawa, Y., Snavely, N., Curless, B., Seitz, S. M., and Szeliski, R., "Reconstructing rome," *Computer*, vol. 43, pp. 40–47, 2010.

[3] Alvarez, L., Deriche, R., Sánchez, J., and Weickert, J., "Dense disparity map estimation respecting image discontinuities : A pde and scale-space based approach," *Journal of Visual Communication and Image Representation*, vol. 13, pp. 3–21, 2002.

[4] Benetazzo, A., "Measurements of short water waves using stereo matched image sequences," *Coastal Engineering*, vol. 53, no. 12, pp. 1013–1032, December 2006.

[5] Boccotti, P., *Wave Mechanics for Ocean Engineering*. Elsevier Oceanography Series, Oxford: Elsevier, 2000.

[6] Bouguet, J. Y., "Camera calibration toolbox for matlab," 2004. Available on-line at http://www.vision.caltech.edu/bouguetj/.

[7] Briggs, W. L., Henson, V. E., and McCormick, S. F., *A Multigrid Tutorial, Second Edition*. SIAM, 2000.

[8] Buzbee, B. L., Golub, G. H., and Nielson, C. W., "On direct methods for solving poisson's equations," *SIAM Journal on Numerical Analysis*, vol. 7, no. 4, pp. 627–656, 1970.

[9] Caprile, B. and Torre, V., "Using vanishing points for camera calibration," *Int. J. Comput. Vision*, vol. 4, no. 2, pp. 127–140, 1990.

[10] Chopp, D., "Computing minimal surfaces via level set curvature flows," *J. Comparative Physics*, vol. 106, no. 1, pp. 77–91, 1993.

[11] Chopp, D. L. and Sethian, J. A., "Flow under curvature: Singularity formation, minimal surfaces, and geodesics," *Journal of Experimental Mathematics*, vol. 2, no. 4, pp. 235–255, 1993.

[12] Clarke, T. and Fryer, J., "The development of camera calibration methods and models," *Photogrammetric Record*, vol. 16, no. 91, pp. 51–66, 1998.

[13] Coté, L., Davis, J., Marks, W., McGough, R., Mehr, E., Pierson, W., Ropek, J., Stephenson, G., and Vetter, R., "The directional spectrum of wind generated sea as determined from data obtained by the stereo wave observation project," *Metereological paper*, vol. 2, no. 6, 1960.

[14] Dugan, J. P., Fetzer, G. J., Bowden, J., Farruggia, G. J., Williams, J. Z., Piotrowski, C. C., Vierra, K., Campion, D., and Sitter, D. N., "Airborne optical system for remote sensing of ocean waves," *Journal of Atmospheric and Oceanic Technology*, vol. 18, no. 7, pp. 1267–1276, 2001.

[15] Dugan, J. P., Piotrowski, C. C., and Williams, J. Z., "Water depth and surface current retrievals from airborne optical measurements of surface gravity wave dispersion," *Journal of Geophysical Research*, vol. 106, pp. 16903–16915, August 2001.

[16] Evans, L. C., *Partial Differential Equations (Graduate Studies in Mathematics, V. 19)*. American Mathematical Society, June 1998.

[17] Farago I., Karatson, J., *Numerical Solution of Nonlinear Elliptic Problems Via Preconditioning Operators: Theory and Applications*. Nova Science Publishers, 2010.

[18] Faugeras, O., Luong, Q.-T., and Papadopoulo, T., *The Geometry of Multiple Images*. The MIT Press, 2001.

[19] Faugeras, O. D. and Keriven, R., "Variational principles, surface evolution, pdes, level set methods, and the stereo problem," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 336–344, 1998.

[20] Fedele, F., Gallego, G., Yezzi, A., and Benetazzo, A., "Understanding extreme waves via a variational wave acquisition stereo system," in *Workshop ROGUE WAVES*, 2008. Oct 13-15, Brest, France.

[21] Fedele, F. and Tayfun, A., "On nonlinear wave groups and crest statistics," *Journal of Fluid Mechanics*, vol. 620, pp. 221–239, 2009.

[22] Fedele, F., "Rogue wave in oceanic turbulence," *PHYSICA D*, vol. 237, no. 14-17, pp. 2127–2131, 2008.

[23] Fedele, F., Gallego, G., Benetazzo, A., Yezzi, A., and Tayfun, M., "Euler characteristics and maxima of oceanic sea states," in *31 Convegno di Idraulica e Costruzioni Idrauliche*, 2008. Sept. 9-12, Perugia Italy.

[24] Fischler, M. A. and Bolles, R. C., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[25] Forsyth, D. A. and Ponce, J., *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.

[26] Fua, P., "A parallel stereo algorithm that produces dense depth maps and preserves image features," *Machine Vision and Applications*, vol. 6, pp. 35–49–49, December 1993.

[27] Furukawa, Y. and Ponce, J., "Accurate camera calibration from multi-view stereo and bundle adjustment," *International Journal of Computer Vision*, vol. 84, no. 3, pp. 257–268, 2009.

[28] Furukawa, Y. and Ponce, J., "Accurate, dense, and robust multiview stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1362–1376, 2010.

[29] Gallego, G., Benetazzo, A., Yezzi, A., and Fedele, F., "Wave statistics and spectra via a variational wave acquisition stereo system," in *ASME 27th International Conference on Offshore Mechanics and Arctic Engineering*, 2008. Lisbon, Portugal.

[30] Gamble, E. and Poggio, T., "Visual integration and detection of discontinuities: The key role of intensity edges," tech. rep., Massachusetts Institute of Technology, Cambridge, MA, USA, 1987.

[31] Gander, W., G, W., and Golub, G. H., "Cyclic reduction - history and applications," in *Workshop on Scientific Computing*, (Hong Kong), 1997.

[32] Golub, G. H., Huang, L. C., Simon, H., and pai Tang, W., "A fast poisson solver for the finite difference solution of the incompressible navier-stokes equations," *SIAM J. Sci. Comput*, vol. 19, pp. 1606–1624, 1998.

[33] Hackbusch, W., *Multi-Grid Methods And Applications)*. Springer, 2009.

[34] Hao, K.-R. and Ding, Y.-S., "Trinocular matching realized by a monocular stereovision sensor for parallel manipulator," in *10th International Conference on Control, Automation, Robotics and Vision - ICARCV*, pp. 1436–1441, 2008.

[35] Hartley, R. I. and Zisserman, A., *Multiple View Geometry in Computer Vision*. Cambridge University Press, second ed., 2004.

[36] Hartley, R. I., "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 580–593, 1997.

[37] Hartley, R. I., "Kruppa's equations derived from the fundamental matrix," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 133–135, 1997.

[38] Hartley, R. I. and Sturm, P. F., "Triangulation," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.

[39] Hauser, D., Kahma, K. K., Krogstad, H. E., Lehner, S., Monbaliu, J., and Wyatt, L. R., *European COST Action 714 (Measuring and Analysing the Directional Spectrum of Ocean Waves)*. European Commission/European Science Fundation, 2003.

[40] Hockney, R. W., "A fast direct solution of poisson's equation using fourier analysis," *Journal of the Association for Computing Machinery*, vol. 12, no. 1, pp. 95–113, 1965.

[41] Holland, K., Holman, R., Lippmann, T., Stanley, J., and Plant, N., "Practical use of video imagery in nearshore oceanographic field studies," *Oceanic Engineering, IEEE Journal of*, vol. 22, pp. 81 –92, jan. 1997.

[42] Holthuijsen, L. H., "Observations of the directional distribution of ocean-wave energy in fetch-limited conditions," *Journal of Physical Oceanography*, vol. 13, no. 2, pp. 191–207, 1983.

[43] Holthuijsen, L., "The directional energy distribution of wind generated waves as inferred from stereographic observations of the sea surface," Tech. Rep. 81-2, Dept. Civ. Eng. Delft University of Technology, 1981.

[44] Hsiao, Y.-H. and Huang, M.-C., "Application of active contour model in tracking sequential nearshore waves," *China Ocean Engineering*, vol. 23, no. 12, pp. 251–266, 2009.

[45] Intel, C., "Open source computer vision library," 1999-2010. Available on-line at http://opencv.willowgarage.com/.

[46] Jin, H., *Variational methods for shape reconstruction in computer vision*. PhD thesis, Washington University, St. Louis, MO, USA, 2003. Director: Soatto, Stefano.

[47] Jin, H., Soatto, S., and Yezzi, A. J., "Stereoscopic shading: Integrating mult1frame shape cues in a variational framework," in *CVPR*, pp. 1169–1176, 2000.

[48] Jin, H., Yezzi, A. J., and Soatto, S., "Variational multiframe stereo in the presence of specular reflections," *3D Data Processing Visualization and Transmission, International Symposium on*, vol. 0, p. 626, 2002.

[49] Jin, H., Yezzi, A. J., Tsai, Y.-H. R., Cheng, L.-T., and Soatto, S., "Estimation of 3d surface shape and smooth radiance from 2d images: A level set approach," *J. Sci. Comput.*, vol. 19, no. 1-3, pp. 267–292, 2003.

[50] Jolliffe, I., *Principal Component Analysis*. Springer Series in Statistics, 2nd ed., Springer, NY, 2002.

[51] Kelley, C. T., "Iterative methods for linear and nonlinear equations," *Frontiers in Applied Mathematics*, vol. SIAM, 1995.

[52] Kelley, C. T., "Iterative methods for optimization," *Frontiers in Applied Mathematics*, vol. SIAM, 1999.

[53] Kitaigorodskii, S. A., Krasitskii, V. P., and Zaslavskii, M. M., "On phillips theory of equilibrium range in the spectra of wind-generated gravity waves," *Journal of Physical Oceanography*, vol. 5, pp. 410–420, 1975.

[54] Liebowitz, D. and Zisserman, A., "Combining scene and auto-calibration constraints," in *ICCV99*, pp. 293–300, 1999.

[55] Liu, P. and MacHutchon, K., "Wave heights in a 4d ocean wave field," in *ASME 27th International Conference on Offshore Mechanics and Arctic Engineering*, 2008. Lisbon, Portugal.

[56] Longuet-Higgins, M. S., "The effect of non-linearities on statistical distributions in the theory of sea waves," *Journal of fluid mechanics*, vol. 17, pp. 459–480, 1963.

[57] Lorensen, W. E. and Cline, H. E., "Marching cubes: A high resolution 3d surface construction algorithm," in *SIGGRAPH*, vol. 21-4, July 1987.

[58] Lu, Y., Zhang, J., Wu, Q., and Li, Z.-N., "A survey of motion-parallax-based 3-d reconstruction algorithms," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 34, pp. 532 –548, nov. 2004.

[59] Ma, Y., Soatto, S., Kosecka, J., and Sastry, S., *An Invitation to 3D Vision: From Images to Geometric Models*. Springer Verlag, 2003.

[60] MacHutchon, K. R., Wessels, W., Wu, C. H., and Liu, P. C., "The use of streamed digital video data and binocular stereoscopic image system (bisis) processing methods to analyze ocean wave field kinematics," in *ASME 28th International Conference on Offshore Mechanics and Arctic Engineering*, 2009. Honolulu, Hawaii.

[61] MacHutchon, K. and Liu, P., "Measurement and analysis of ocean wave fields in four dimensions," in *ASME 26th International Conference on Offshore Mechanics and Arctic Engineering*, vol. 1, pp. 923–927, 2007.

[62] Maybank, S. J. and Faugeras, O. D., "A theory of self-calibration of a moving camera," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 123–151, 1992.

[63] McCormick, S. F. E., "Multigrid methods," *Frontiers in Applied Mathematics*, 1987.

[64] Menard, C., "Stereo correspondences in scale space," in *Computer Vision ACCV'98* (Chin, R. and Pong, T.-C., eds.), vol. 1351 of *Lecture Notes in Computer Science*, pp. 550–557, Springer Berlin / Heidelberg, 1997.

[65] Microsoft, "Photosynth," 2006-2010. http://photosynth.net.

[66] Mumford, D. and Shah, J., "Optimal approximations by piecewise smooth functions and variational problems," *Comm. Pure Appl. Math*, vol. XLII, no. 5, pp. 577–685, 1988.

[67] N.F., B., "Finding the direction of travel of sea waves," *Nature*, vol. 154, p. 10481049, 1954.

[68] Orszag, S. A., "Numerical methods for the simulation of turbulence," *Phys. Fluids Supp. II*, vol. 12, pp. 250–257, 1969.

[69] Osher, S. and Sethian, J. A., "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[70] Paragios, N. K. and Deriche, R., "A pde-based level-set approach for detection and tracking of moving objects," in *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, (Washington, DC, USA), p. 1139, IEEE Computer Society, 1998.

[71] Pearson, K., "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559–572, 1901.

[72] Pollefeys, M., *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, Katholieke Universiteit Leuven - Faculteit Toegepaste Wetenschappen Arenbergkasteel, B-3001 Heverlee (Belgium), May 1999.

[73] Pollefeys, M., Gool, L. J. V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., and Koch, R., "Visual modeling with a hand-held camera," *International Journal of Computer Vision*, vol. 59, no. 3, pp. 207–232, 2004.

[74] Pollefeys, M., Koch, R., Vergauwen, M., and Gool, L. J. V., "Flexible 3d acquisition with a monocular camera," in *ICRA*, pp. 2771–2776, 1998.

[75] Ponce, J., McHenry, K., Papadopoulo, T., Teillaud, M., and Triggs, B., "On the absolute quadratic complex and its application to autocalibration," in *CVPR (1)*, pp. 780–787, IEEE Computer Society, 2005.

[76] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., *Numerical Recipes: The Art of Scientific Computing, Third Edition*. Cambridge University Press, 2007.

[77] Rocha, K., *A variational approach for viewpoint-based visibility maximization*. PhD thesis, Georgia Institute of Technology, Atlanta, GA, USA, 2008. Director: Yezzi, Anthony.

[78] Ronda, J. I., Valdés, A., and Gallego, G., "Line geometry and camera autocalibration," *J. Math. Imaging Vis.*, vol. 32, no. 2, pp. 193–214, 2008.

[79] Ronda, J. I., Valdés, A., and Jaureguizar, F., "Camera autocalibration and horopter curves," *International Journal of Computer Vision*, vol. 57, no. 3, pp. 219–232, 2004.

[80] Saad, Y., *Iterative Methods for Sparse Linear Systems, Second Edition*. SIAM, 2003.

[81] Santel, F., Linder, W., and Heipke, C., "Stereoscopic 3d-image sequence analysis of sea surfaces," *Proceedings of the ISPRS Commission V Symposium*, vol. 35, no. 5, pp. 708–712, 2004.

[82] Scharstein, D. and Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2002. 10.1023/A:1014573219977.

[83] Schmid, P. J., "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, July 2010.

[84] Schumacher, A., "Stereophotogrammetrische wellenaufnahmen. wiss ergeb. dtsch. atlant. exped. forschungs vermessung. meteor 1925-1927," *Ozeanographische Sonderuntersuchungen, Erste Lieferung,*, 1939.

[85] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R., "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, (Washington, DC, USA), pp. 519–528, IEEE Computer Society, June 2006.

[86] Seitz, S. M. and Dyer, C. R., "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision*, vol. 35, pp. 1067–1073, 1997.

[87] Sethian, J. A., *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.

[88] Shapiro, L. G., Stockman, G. C., Shapiro, L. G., and Stockman, G., *Computer Vision*. Prentice Hall, January 2001.

[89] Shemdin, O., "Measurement of short surface waves with stereophotography," in *OCEANS '90. 'Engineering in the Ocean Environment'. Conference Proceedings*, pp. 568 –571, sep. 1990.

[90] Shemdin, O. and Tran, H., "Measuring short surface waves with stereography," *Photogrammetric Engineering and Remote Sensing*, vol. 58, no. 311-316, 1992.

[91] Shemdin, O., Tran, H., and Wu, S., "Directional measurements of short ocean waves with stereography," *J. Geophys. Res.*, vol. 93, pp. 13891–13901, 1988.

[92] Snavely, N., Seitz, S. M., and Szeliski, R., "Modeling the world from internet photo collections," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 189–210, 2008.

[93] Soatto, S., Yezzi, A. J., and Jin, H., "Tales of shape and radiance in multi-view stereo," in *ICCV*, pp. 974–981, 2003.

[94] Socquet-Juglard, H., Dysthe, K., Trulsen, K., Krogstad, H. E., and Liu, J., "Probability distributions of surface gravity waves during spectral changes," *Journal of Fluid Mechanics*, vol. 542, pp. 195–216, 2005.

[95] Stagonas, D. and Muller, G., "Wave field mapping with particle image velocimetry," *Ocean Engineering*, vol. 34, no. 11-12, pp. 1781–1785, 2007.

[96] Strang, G. and Fix, G., *An Analysis of The Finite Element Method*. Prentice Hall, 1973.

[97] Sugimori, Y., "A study of the application of the holographic method to the determination of the directional spectrum of ocean waves," *Deep Sea Research and Oceanographic Abstracts*, vol. 22, no. 5, pp. 339 – 342, IN17–IN23, 343–350, 1975.

[98] Tayfun, A. and Fedele, F., "Wave height distributions and nonlinear effects," *Ocean Engineering*, vol. 34, no. 11-12, pp. 1631,1649, 2007.

[99] Thomas, J., *Numerical Partial Differential Equations: Finite Difference Methods*. Springer, 2010.

[100] Tran, H. M. and Shemdin, O., "Automatic digital correlation technique in stereophotography for measuring short surface waves in the ocean," in *Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing., 1989 International*, vol. 3, pp. 1504 –1504, jul. 1989.

[101] Trefethen, L. N., *Spectral methods in MATLAB*. Philadelphia, PA: SIAM: Society for Industrial and Applied Mathematics, 2000.

[102] Triggs, B., "Autocalibration and the absolute quadric," in *CVPR*, pp. 609–614, IEEE Computer Society, 1997.

[103] Trottenberg, U., *Multigrid*. Academic Press, 2000.

[104] Trucco, E. and Verri, A., *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[105] Tsai, R. Y., "A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE J. Robotics Automat.*, vol. RA-3, no. 4, pp. 323–344, 1987.

[106] Vogiatzis, G., Esteban, C. H., Torr, P. H. S., and Cipolla, R., "Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 2241–2246, 2007.

[107] Wanek, J. M. and Wu, C. H., "Automated trinocular stereo imaging system for three-dimensional surface wave measurements," *Ocean Engineering*, vol. 33, no. 5-6, pp. 723–747, April 2006.

[108] Weickert, J., Ishikawa, S., and Imiya, A., "Linear scale-space has first been proposed in japan," *Journal of Mathematical Imaging and Vision*, vol. 10, pp. 237–252, 1999. 10.1023/A:1008344623873.

[109] Wesseling, P., *An Introduction to Multigrid Methods*. R.T. Edwards, Inc., 2004.

[110] Yezzi, A. and Soatto, S., "Stereoscopic segmentation," *International Journal of Computer Vision*, vol. 53, pp. 31–43, June 2003.

[111] Yuille, A. and Poggio, T., "A generalized ordering constraint for stereo correspondence," in *MIT AI Memo*, 1984.

[112] Zakharov, V. E., "Statistical theory of gravity and capillary waves on the surface of a finite-depth fluid," *Eur. J. Mech. B - Fluids*, vol. 18, no. 3, pp. 327–344, 1999.

[113] Zhang, Z., "Determining the epipolar geometry and its uncertainty: A review," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 161–198, 1998.

[114] Zhang, Z., "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

VITA

Guillermo Gallego Bonet was born in Ávila (Spain) in 1980. He majored in Telecommunications Engineering in 2004 at the E.T.S. Ingenieros de Telecomunicación of the Universidad Politécnica de Madrid, Spain. In July 2005 he earned a Master's degree in Mathematical Engineering (Magíster en Ingeniería Matemática) from the Universidad Complutense de Madrid. In 2005, he was a recipient of the Fulbright Scholarship to pursue graduate studies in the U.S.A. At The Georgia Institute of Technology, he obtained an M.S. in Electrical and Computer Engineering in 2007, an M.S. in Mathematics in 2009 and, in December 2010, the Ph.D. in Electrical and Computer Engineering. His research interests lie within the fields of applied image processing and computer vision.